



**FEATURE SELECTION AND CLASSIFIER DEVELOPMENT FOR RADIO
FREQUENCY DEVICE IDENTIFICATION**

DISSERTATION

Trevor J. Bihl, Civilian

AFIT-ENG-DS-15-D-003

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT-ENG-DS-15-D-003

FEATURE SELECTION AND CLASSIFIER DEVELOPMENT FOR RADIO
FREQUENCY DEVICE IDENTIFICATION

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Trevor J. Bihl, BS MS

Civilian

24 December 2015

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

AFIT-ENG-DS-15-D-003

FEATURE SELECTION AND CLASSIFIER DEVELOPMENT FOR RADIO
FREQUENCY DEVICE IDENTIFICATION

DISSERTATION

Trevor J. Bihl, BS MS

Civilian

Committee Membership:

Michael A. Temple, PhD
Chair

Kenneth W. Bauer Jr., PhD
Member

Robert F. Mills, PhD
Member

ADEDEJI B. BADIRU, Ph.D.
Dean, Graduate School of Engineering
and Management

Abstract

The proliferation of simple and low-cost devices, such as IEEE 802.15.4 “ZigBee” and Z-Wave, in CriticalInfrastructure (CI) increases security concerns. Radio Frequency “Distinct Native Attribute” (RF-DNA) Fingerprinting facilitates biometric-like identification of electronic devices emissions from variances in device hardware. Developing reliable classifier models using RF-DNA fingerprints is thus important for device discrimination to enable reliable *Device Classification* (a one-to-many looks “most like” assessment) and *Device ID Verification* (a one-to-one looks “how much like” assessment). AFIT’s prior RF-DNA work focused on Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) and Generalized Relevance Learning Vector Quantized Improved (GRLVQI) classifiers. This work 1) introduces a new GRLVQI-Distance (GRLVQI-D) classifier that extends prior GRLVQI work by supporting alternative distance measures, 2) formalizes a framework for selecting competing distance measures for GRLVQI-D, 3) introducing response surface methods for optimizing GRLVQI and GRLVQI-D algorithm settings, 4) develops an MDA-based Loadings Fusion (MLF) Dimensional Reduction Analysis (DRA) method for improved classifier-based feature selection, 5) introduces the F-test as a DRA method for RF-DNA fingerprints, 6) provides a phenomenological understanding of test statistics and p-values, with KS-test and F-test statistic values being superior to p-values for DRA, and 7) introduces quantitative dimensionality assessment methods for DRA subset selection.

The optimized GRLVQI algorithm and the proposed GRLVQI-D algorithm show improved performance over the baseline GRLVQI algorithm. When considering the optimized GRLVQI and GRLVQI-D classifiers using $N_F = 189$ Z-Wave features and an arbitrary average correct classification (%C) of %C = 90% benchmark, demonstrated *Device Classification SNR* gain (G_{SNR}) performance relative to baseline GRLVQI includes 1) improved $G_{SNR} = +1.84$ dB using GRLVQI-D with a Cosine distance measure, and 2) best case $G_{SNR} = +1.94$ dB using the GRVLQI optimized algorithm. For *Z-Wave Device ID Verification*, results of included correct verification of *authorized* device IDs (% V_A) include 1) worst case % $V_A = 33.33\%$ for baseline GRLVQI, 2) improved % $V_A = 66.66\%$ for GRLVQI-D using a Cosine distance measure, and 3) best case % $V_A = 100\%$ using the optimized GRLVQI algorithm.

The proposed F-test and MLF DRA methods are shown to offer distinct performance improvements. *ZigBee Device Classification* results for selected DRA methods with an MDA/ML classifier benchmark of %C = 90%, included *SNR* gain relative to the benchmark GRLVQI DRA with $N_{DRA} = 50$ feature sets of 1) $G_{SNR} = +0.82$ dB for MLF DRA, and 2) $G_{SNR} = +0.10$ dB for F-test DRA using $N_{DRA} = 50$. *ZigBee Device ID Verification* results, using the same $N_{DRA} = 50$ feature sets and MDA/ML classifier, included correct % V_A and correct detection of unauthorized *rogue* device IDs (% V_R) of % $V_A = 50\%$ and % $V_R = 91.67\%$ for the benchmark GRLVQI DRA, with 1) comparable % $V_A = 50\%$ and % $V_R = 91.67\%$ for MLF DRA, and 2) best case % $V_A = 75\%$ and % $V_R = 91.67\%$ for F-test DRA.

AFIT-ENG-DS-15-D-003

To the memory of my father.

Acknowledgements

I owe a large debt of gratitude to both my primary research advisor, Dr. Michael Temple, and my secondary research advisor, Dr. Kenneth Bauer, for their patience, guidance, and support throughout this research effort. The six plus years I have worked for and learned from Dr. Bauer, have been a phenomenal joy where I have greatly expanded in my knowledge and capabilities. Additionally, my committee, Dr. Robert Mills and Maj. Brian Stone PhD, also deserve mention for their patience and support in this process.

Particular thanks goes to my fellow AFIT comrade Maj. Todd Paciencia PhD, who provided insight, expertise, camaraderie, collaboration, math help, and explanations of various esoteric (to me at least) concepts, and Dr. James Cordeiro who provided needful guidance at opportune times. My office mates in the Sensor Fusion Lab during this ordeal provided excellent additional support: Capt. Michael Gibb, with his cheerful attitude, and Maj. Joseph Bellucci, for being a source of his constant level-headedness.

Finally, the lifelong support and encouragement of my parents have made this possible and I will forever be grateful for that. My wife, the real Dr. Bihl, deserves particular thanks for her love, patience, constant encouragement, understanding, and support during this ordeal. Additionally, special mention needs to go to my good friends caffeine and coffee, without whom nothing would have been possible.

Trevor J. Bihl

Table of Contents

	Page
Abstract	iv
Dedication	vi
Acknowledgements	vii
Table of Contents	viii
List of Figures	xii
List of Tables	xii
I. Introduction	21
1.1 Operational Motivation	22
1.2 Radio Frequency Fingerprinting	24
1.3 Technical Motivation	26
1.4 Research Contribution	27
1.5 Document Organization	28
II. Background	30
2.1 Introduction	30
2.2 Signals of Interest: Wireless Networks	30
2.2.1 IEEE 802.15.4 ZigBee Devices	32
2.2.2 IEEE 802.15.4 Z-Wave Devices	37
2.2.3 Post-Collection Data Manipulation	39
2.3 Physical Layer Device Identification	39
2.3.1 RF Device Emissions	40
2.4 1D Time Domain (TD) RF-DNA Fingerprints	44
2.4.1 ZigBee and Z-Wave RF-DNA Fingerprinting	47
III. Statistical Pattern Recognition	49
3.1 Introduction	50
3.2 Multiple Discriminant Analysis	51
3.2.1 MDA Feature Relevance Ranking	54
3.2.2 Maximum Likelihood (ML) Device Classification	56
3.3 Learning Vector Quantization Family of Methods	57

3.3.1 Gradient Descents and LVQ	60
3.4 Device Classification and Verification Methodology.....	79
3.4.1 Classification Performance	79
3.4.2 Device ID Verification.....	83
3.4.3 MDA/ML and GRLVQI Baseline Results.....	86
3.4.4 MDA versus GRLVQI in RF-DNA Research	89
IV. Dimensional Reduction Analysis	91
4.1 Introduction.....	91
4.2 Dimensional Reduction Analysis Methods.....	92
4.2.1 Distribution Based Feature Selection DRA	95
4.2.2 Post-Classification Feature Selection DRA for RF-DNA	116
4.2.3 Algorithmic Fusion Methods	122
4.2.4 DRA Fusion Methods	130
4.2.5 Random Feature Selection	134
4.2.6 Dimensionality Assessment	135
4.3 DRA Applications to ZigBee Data	141
4.3.1 DRA Method Comparisons.....	141
4.3.2 DRA Method Classification Performance Assessments.....	147
4.3.3 DRA Method Verification Performance Assessments	157
V. Extensions to the LVQ-Family of Algorithms.....	160
5.1 Introduction.....	160
5.2 GRLVQI-D Algorithm Development	162
5.2.1 Prior Implementations of non-Euclidean Distances in LVQ	162
5.2.2 Developing a Differentiation Skeleton for LVQ Improvements	164
5.3 Selecting Distance Measures for GRLVQI-D	173
5.3.1 Selecting Distance Measures for Consideration	175
5.3.2 Comparing Potential Distance Metrics via Correlation	175
5.3.3 Determining Suitable Distance Measures and LVQ Algorithm Settings.....	180
5.3.4 Relevance Learning with Alternative LVQ Distance Measures.	183
5.3.5 Distance Measure Extensions to GRLVQ and GRLVQI	186
5.4 GRVLQI-D Extension for RF-DNA Fingerprinting.....	189
5.4.1 LVQ Architecture Selection and Specification.....	190

5.4.2 Experimental Design for GRLVQI-D Algorithmic Settings	203
5.4.3 GRLVQI-D Performance Results	209
5.4.4 Results Interpretation	217
VI. Improvements to the RF-DNA Fingerprinting Process	219
6.1 Introduction	219
6.2 Normalization, Standardization and Phase Feature Dominance	220
6.2.1 Phenomenology of Amplitude, Frequency, and Phase	220
6.2.2 Normalization and Standardization	221
6.3 Simulation Methods, Dependence and Correlation Effects in RF-DNA	225
6.3.1 Transient Determination	227
6.3.2 Autocorrelation and the Number of Batches	228
VII. Summary and Conclusions	232
7.1 Research Summary	232
7.2 Research Contributions	235
7.3 Proposed Future Research	237
7.4 Sponsor Acknowledgement	239
APPENDIX A: Lemma Associated with Multiple Discriminant Analysis Loadings	240
APPENDIX B: Examination of LVQ and GLVQ Properties and Features	241
APPENDIX C: <i>P</i> -values versus Test Statistics on Selected Academic Datasets	244
APPENDIX D: DRA Method Fusion Classification and Verification Performance Assessments	254
APPENDIX E: Gradient Descent and Derivatives in GLVQ Family Algorithms	257
APPENDIX F: Gradient Descent in GRLVQ and GRLVQI Relevance Computation ..	263
APPENDIX G: Cost Function Extensions for the GLVQ Family of Algorithms	266
APPENDIX H: Relevance Derivatives for GRLVQI	269
APPENDIX I: Review of Distance Measures	272
APPENDIX J: Derivatives a LVQ nd Prototype Vectors Updates for Selected Distance Metrics	282

APPENDIX K: Design of Experiments Results	290
APPENDIX L: MEX File Programming Considerations	296
APPENDIX M: GRLVQI-D Performance on ZigBee RF-DNA Fingerprints with Z-Wave Based Optimization.....	298
VIII. Bibliography.....	306
Vita.....	360

List of Figures

	Page
Figure II-1: General operations of digital communication, adapted from [64, 138].	31
Figure II-2: ZigBee PHY layer packet structure, adapted from [167].	34
Figure II-3: Conceptualization of ZigBee data collection, from [91].	35
Figure II-4: Z-Wave PHY layer packet structure, adapted from [173].	38
Figure II-5: RF-DNA Fingerprinting Architecture, adapted from Cobb et al. [19].	43
Figure II-6: Traditional RF-DNA Feature Extraction Approach as Applied to ZigBee Devices, adapted from [91].	44
Figure III-1: Conceptualization of a) MDA class projections from [216] and b) LVQ prototype development as adapted from [51, 216].	51
Figure III-2: General taxonomy of ANN approaches, adapted from [254] using the ANN families of [250, p. 368].	58
Figure III-3: Conceptualization of the differences between a) ANNs and b) LVQ networks, adapted from [250, 262].	59
Figure III-4: LVQ prototype vector update conceptualization; adapted from [249].	60
Figure III-5: Conceptualization of the LVQ2 and LVQ2.1 prototype vector update approach using the LVQ2.1 process; adapted from [282].	66
Figure III-6: ZigBee Full-dimensional Baseline Classification Results	80
Figure III-7: Z-Wave Full-dimensional Baseline Classification Results	80
Figure III-8: Gain Trade Off Example for MDA/ML (TST) and GRLVQI (TST) for ZigBee.	82
Figure III-9: ZigBee MDA/ML and GRLVQI full-dimensionality authorized device verification results baseline	85
Figure III-10: ZigBee MDA/ML and GRLVQI full-dimensionality rogue device verification results baseline	85
Figure III-11: Z-Wave MDA/ML and GRLVQI full-dimensionality authorized device verification results baseline	86

Figure IV-1: p-value KS-test Feature Selection Algorithm as adapted from Patel [134] and modified herein.....	97
Figure IV-2: Sum of p -values from pairwise KS-test on ZigBee training observations using a full-dimensional ($N_F = 729$) feature set at $SNR = 10$ dB [89, 113]. Lower values indicate greater feature difference and potentially greater relevance.....	98
Figure IV-3: Mean of test statistic values from pairwise KS-test on ZigBee training observations using a full-dimensional ($N_F = 729$) feature set at $SNR = 10$ dB. Higher values indicate more different (and possibly more relevant) features.	99
Figure IV-4: Example p -value computation from test statistics using an F-distribution.	102
Figure IV-5: One way ANOVA F-test Feature Relevance Ranking Algorithm.....	102
Figure IV-6: Test statistic values from F-test on ZigBee training observations using a full-dimensional ($N_F = 729$) feature set at $SNR = 10$ dB. Lower values indicate greater feature difference and potentially greater relevance.....	103
Figure IV-7: p -values from F-test on ZigBee training observations using a full-dimensional ($N_F = 729$) feature set at $SNR = 10$ dB. Lower values indicate greater feature difference and potentially greater relevance.	104
Figure IV-8: Normalized histogram of <i>summed</i> pairwise KS-test p -values using a full-dimensional ($N_F = 729$) ZigBee TNG feature set for varying $SNR = [0, 10, 18, 30]$ dB.....	110
Figure IV-9: Normalized histograms of <i>mean</i> pairwise KS-test statistic values using a full-dimensional ($N_F = 729$) ZigBee TNG feature set for $SNR = [0, 10, 18, 30]$ dB.....	111
Figure IV-10: Normalized histogram of F-test p -values using a full-dimensional ($N_F = 729$) ZigBee TNG feature set for varying $SNR = [0, 10, 18, 30]$ dB.....	112
Figure IV-11: Normalized histograms of F -test statistic values using a full-dimensional ($N_F = 729$) ZigBee TNG features for $SNR = [0, 10, 18, 30]$ dB.....	113
Figure IV-12: Feature ranking using GRLVQI relevance values using full-dimensional $N_F = 729$ ZigBee TNG observations at $SNR = 10$ dB.	117
Figure IV-13: Feature ranking values from Wilk's Lambda ratio using full-dimensional $N_F = 729$ ZigBee TNG observation at $SNR = 10$ dB.....	119

Figure IV-14: ZigBee discriminant loadings (L) for the three discriminant functions using full-dimensional $N_F = 729$ ZigBee TNG observations at $SNR = 10$ dB. Reprinted from [135].	121
Figure IV-15: Three General Fusion Method Architectures, adapted from [418].	124
Figure IV-16: Feature ranking values from Unscaled Maximum ($UMax$) discriminant loadings using full-dimensional $N_F = 729$ ZigBee TNG observations at $SNR = 10$ dB.	127
Figure IV-17: Matrix scatterplots for four MDA Loadings Fusion (MLF) methods, Unscaled ($UMax$ and $USum$) and Scaled ($SMax$ and $SSum$), using full-dimensional $N_F = 729$ feature set at $SNR = 10$ dB. Reprinted from [135].	130
Figure IV-18: Generic Example of Score and Rank Fusion	131
Figure IV-19: General Process for Concatenation Fusion	134
Figure IV-20: Top ranked $N_F = 10$ reduced dimensional feature sets by DRA method, reprinted from [135].	144
Figure IV-21: Top ranked $N_F = 26$ reduced dimensional feature sets by DRA method.	145
Figure IV-22: Top ranked $N_F = 50$ reduced dimensional feature sets by DRA method.	146
Figure IV-23: ZigBee MDA/ML <i>Testing (TST)</i> classification performance for $N_{DRA} = 10$ reduced dimensional feature sets, reprinted from [135].	149
Figure IV-24: ZigBee MDA/ML <i>Testing (TST)</i> classification performance for $N_{DRA} = 26$ reduced dimensional feature sets, reprinted from [135].	150
Figure IV-25: ZigBee MDA/ML <i>Testing (TST)</i> Classification performance for $N_{DRA} = 50$ reduced dimensional feature sets, reprinted from [135].	151
Figure IV-26: ZigBee MDA/ML <i>Testing (TST)</i> classification performance for each DRA method at $SNR = 10$ dB. $N_{DRA} = [10, 26, 50, 100, 157, 191, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700]$ reduced dimensional feature sets are evaluated to understand how DRA fundamentally impacts performance. Reprinted from [135].	152
Figure IV-27: ZigBee <i>Device ID Verification</i> performance for the $N_{DRA} = 50$ $UMax$ feature subset at $SNR = 10$ dB. Reprinted from [135].	157
Figure V-1: Components of GLVQ, GRLVQ and GRLVQI Gradient Descents.	171

Figure V-2: Components of GLVQ, GRLVQ and GRLVQI Gradient Descents.	171
Figure V-3: Pseudocode Process and Derivative Skeleton for Changing Distance Metrics in LVQ, RLVQ, GLVQ, and GRLVQ.	172
Figure V-4: Iterative Process for Selecting Distance Metrics for GRLVQI.....	174
Figure V-5: Correlation Comparison of Distance Metrics on Random Normal Data. ...	177
Figure V-6: Dendrogram with Complete Linkage and Correlation Matrix, from Figure V-5, as Distance Matrix.....	179
Figure V-7: Distance Measure Performance versus Learning Rate for LVQ.....	182
Figure V-8: Distance Measure Performance versus Learning Rate for GLVQ.....	183
Figure V-9: Learning Rate vs Relevance Learning Rate for Squared Euclidean RLVQ	185
Figure V-10: Learning Rate vs Relevance Learning Rate for Canberra RLVQ.....	186
Figure V-11: Learning Rate vs Relevance Learning Rate for Cosine RLVQ	186
Figure V-12: Learning Rate vs Relevance Learning Rate for Squared Euclidean GRLVQ.	188
Figure V-13: Learning Rate vs Relevance Learning Rate for Canberra GRLVQ.....	188
Figure V-14: Learning Rate vs Relevance Learning Rate for Cosine GRLVQ.	189
Figure V-15: GRLVQI Classification Results on ZigBee RF-DNA Fingerprints at 14 dB Using Various PVs/class.	198
Figure V-16: GRLVQI Classification Performance with 10 PVs versus 13 PVs.....	199
Figure V-17: Verification Performance in GRLVQI with 13 PVs at 8dB.	200
Figure V-18: Verification Performance in GRLVQI with 13 PVs at 14dB.	200
Figure V-19: Verification Performance in GRLVQI with 13 PVs at 18dB.	201
Figure V-20: GRLVQI Classification Performance Using Squared Euclidean Distance Using Optimized Algorithmic Settings.	210
Figure V-21: GRLVQI ID Verification Performance in Squared Euclidean GRLVQI using Optimization Settings at 20dB for Z-Wave Dataset.	212

Figure V-22: GRLVQI-D Classification Performance Using Canberra Distance Using Optimized Algorithmic Settings.....	213
Figure V-23: GRLVQI-D ID Verification Performance in Canberra GRLVQI using Optimization Settings at 20dB for Z-Wave Dataset.....	214
Figure V-24: GRLVQI Classification Performance Using Cosine Distance Using Optimized Algorithmic Settings.....	215
Figure V-25: GRLVQI ID Verification Performance in Cosine GRLVQI using Optimization Settings at 20dB for Z-Wave Dataset.....	216
Figure VI-1: Boxplot of ZigBee RF-DNA Features at $SNR = 10$ dB for Authorized Devices Using the Nominal Mean Centering and Maximum Scaled Normalization process of [18, 19].....	222
Figure VI-2: Boxplots of ZigBee RF-DNA Features at 10dB for Authorized Devices Using Standardized Data.	224
Figure VI-3: General Batching Method for Simulation Output Showing the Response Divided into M Total Batches [243].....	227
Figure VI-4: Autocorrelation of ZigBee Data Features, $SNR = 10$ dB.	231
Figure B-1: Hypothetical Situation with Two PVs and One Exemplar.....	241
Figure B-2: Distances Between Exemplars and a) PV_1 and b) PV_2	242
Figure B-3: General Relationship Between Distance Difference Measure and PV Distances	243
Figure B-4: Relationship Between Distance Difference Measure and PV Distances for Closely Spaced PV_1 and PV_2	243
Figure M-1: ZigBee GRLVQI Classification Performance Using Squared Euclidean Distance Using Optimized Algorithmic Settings.	299
Figure M-2: GRLVQI ID Verification Performance of ZigBee in Squared Euclidean GRLVQI using Z-Wave Determined Classification-Based Optimization Settings at 18dB.....	300
Figure M-3: GRLVQI ID Verification Performance of ZigBee in Squared Euclidean GRLVQI using Z-Wave Determined Vefication-Based Optimization Settings at 18dB.....	301

Figure M-4: GRLVQI Classification Performance Using Cosine Distance Using Optimized Algorithmic Settings.....	302
Figure M-5: GRLVQI ID Verification Performance of ZigBee in Cosine GRLVQI using Z-Wave Determined Classification–Based Optimization Settings at 18dB.	303
Figure M-6: GRLVQI ID Verification Performance of ZigBee in Cosine GRLVQI using Z-Wave Determined Verification–Based Optimization Settings at 18dB. .	304

List of Tables

	Page
Table I-1: OSI Model, adapted from [62–65].	24
Table I-2: Relational mapping between technical contributions in previous related work and <i>current</i> research contributions. The × symbol denotes areas addressed. .	28
Table II-1: ZigBee SHR Region Format, adapted from [91, 167, 168].	35
Table II-2: ZigBee Collected Data, adapted from [91].	36
Table II-3: ZigBee versus Z-Wave, adapted from [170, 172].	37
Table III-1: Major Variations in LVQ Family of Algorithms.	62
Table III-2: Improved GRLVQ Update Rule of Mendenhall [244]	72
Table III-3: Nominal GRLVQ and GRLVQI Learning Parameter Learn Schedule.....	76
Table III-4: Baseline Classification Results.	87
Table III-5: ZigBee Baseline Device ID Verification Results.....	88
Table III-6: Z-Wave Baseline Device ID Verification Results.	89
Table IV-1: Quantity of ZigBee p -values Less Than or Equal to 64-bit Relative Spacing, from [49].	114
Table IV-2: P -values vs Test Statistic for ZigBee at SNR = 10 dB Ordered by Decreasing F-test and KS-Test Statistic Value, adapted from [49]	115
Table IV-3: Dimensionality Assessment by p -value and Significance Level, Reprinted from [49].	138
Table IV-4: Correlation Matrix for DRA Method Scores at SNR = 10 dB, from Bihl et al. [135]. High correlations (>0.8) and low correlations (<0.2) are in bold and shaded light grey.	143
Table IV-5: DRA Subset Statistics for F-test, KS-test, GRLVQI, Wilk’s Lambda, $USum$, $UMax$, $SSum$, and $SMax$. Reprinted from [135].	147
Table IV-6: Relative DRA “Gain” (dB) Over Baseline Performance for $\%C = 90\%$ Classification Accuracy. Bold entries with light grey shading denote best case (lowest gain) performance and bold entries denote values within 10% of the best. Reprinted from [135].	154

Table IV-7: Relative Accuracy Percentage (RAP) from Baseline $N_{DRA} = 729$ Feature Set. Bold entries with light grey shading denote best case (highest scoring) performance. Reprinted from [135].	156
Table IV-8: Device ID Verification Performance For $\%C = 90\%$ at $SNR = 10$ dB: True Verification Rate (TVR) for $N_{Auth} = 4$ Authorized Devices and Rogue Rejection Rate (RRR) For $N_{Auth} \times N_{Rog} = 36$ rogue scenarios. Bold entries denote values within 10% of the Best, and bold entries with light grey shading denote best case performance and. Reprinted from [135].	159
Table V-1: Learning and Relevance Rates for LVQ Algorithm Experiment.	181
Table V-2: Example of ANN Architecture Effects on ANN Performance, reproduced from [493].	191
Table V-3: Example of PV Architecture Effects on LVQ Performance on Insects.	192
Table V-4: #PVs for RF-DNA Using Various Heuristics for ZigBee Data.	196
Table V-5: Relationship between PVs and Classification/Verification Performance	202
Table V-6: Experimental Design Region for GRLVQI.	204
Table V-7: Overview of Second Order Models.	206
Table V-8: Features Significant Per Model.	207
Table V-9: Optimized Algorithms Settings for Z-Wave Data.	209
Table V-10: Z-Wave Optimized Algorithms Results for Z-Wave Data.	217
Table C-1: Example Academic Datasets.	245
Table C-2: p-values vs Test Statistic for Fisher Iris	246
Table C-3: p-values vs Test Statistic for Insect.	247
Table C-4: p-values vs Test Statistic for Vertebral Column.	248
Table C-5: p-values vs Test Statistic for Wine Quality.	248
Table C-6: p-values vs Test Statistic for Wisconsin Breast Cancer.	249
Table C-7: p-values vs Test Statistic for Wine	250

Table C-8: p-values vs Test Statistic for MNIST.	251
Table C-9: p-values vs Test Statistic for Ecoli.	252
Table D-1: Relative DRA “Gain” (dB) Over Baseline Performance for $\%C = 90\%$ Classification Accuracy for DRA Fusion Methods. Bold entries denote values within 10% of the Best, and bold entries with light grey shading denote best case performance and.....	255
Table D-2: Device ID Verification Performance for $\%C = 90\%$ at $SNR = 10$ dB: True Verification Rate (TVR) for $N_{Auth} = 4$ <i>Authorized</i> Devices and <i>Rogue</i> Rejection Rate (RRR) For $N_{Auth} \times N_{Rog} = 36$ rogue scenarios. Bold entries denote values within 10% of the Best, and bold entries with light grey shading denote best case performance and.....	256
Table K-1: Design of Experiments Results	290
Table M-1: GRLVQI Performance for ZigBee RF-DNA Data Using Z-Wave Optimized Algorithmic Settings.	305

I. Introduction

But in war, as in life generally, all parts of the whole are interconnected and thus the effects produced, however small their cause, must influence all subsequent military operations...

—CARL VON CLAUSEWITZ, 1780 – 1831

Communication networks permeate society through commercial networks, such as the internet, cell phones and Wi-Fi, to Industrial Control Systems (ICS), such as Supervisory Control And Data Acquisition (SCADA) systems, which monitor and control many critical infrastructure (CI) systems. In all communication networks, one is interested in a balance between attributes such as performance, security, reliability, availability, and survivability [1, 2]. In CI applications, all of these attributes are necessary since CI interruption can threaten lives, disable governments, affect the economy, and damage ecological systems [3]. Additionally, the “fog of war” has been reduced due to advances in digital communications [4]; however, security concerns can both limit user confidence in communications networks [5] and reduce this functionality [4].

Security is a critical component in communication networks and, due to functional interconnectedness, compromising one point can compromise overall system security [6]. Therefore, the security of communication and industrial networks and devices is of high importance to the Department of Defense [7–9]. Various issues exist in securing hardware [10], including: 1) identifying counterfeited or reused components

[11–17], 2) determining claimed device identities [18, 19], and 3) determining aging effects [20–34].

Improving methods for vetting communication device identity by examining and characterizing device physical properties are of interest. AFIT’s Radio Frequency (RF) Fingerprinting process, RF Distinct Native Attribute (RF-DNA) Fingerprinting [19], is a systematic and proven method for extracting statistical features from waveform data. Of interest in this research was the extension and improvement of RF-DNA practices for improved communication device identification and security.

1.1 Operational Motivation

The “Internet of Things” is predicted to enable wide connectivity between commercial, industrial and consumer devices [35]. However, such connectivity includes many risks due to the possibility of hackers disrupting services, stealing information, or taking control of various devices in CI applications or consumer use [35, 36]. Facilitating the “Internet of Things” is the proliferation of low cost networks, such as those created by IEEE 802.15.4 “ZigBee” and Z-Wave devices, into CI applications present numerous security issues [37, 38].

Both ZigBee and Z-Wave devices have numerous operating advantages that motivate their use in CI applications, such as the ability to communicate up to 100 meters and the ability to sustain networks comprised of up to 65,000 devices [39]. Given these advantages, ZigBee devices are believed to provide interconnections between more physical devices in the world than any other wireless technology [37]. CI networks

frequently include many low cost communication devices, such as ZigBee and Z-Wave, for interaction with physical objects, e.g. power relays [40–43], patient monitoring devices [44], security systems [45], automation and control systems [46], home automation [47], and electric metering [48].

Due to the ubiquity of ZigBee and Z-Wave devices, general security concerns exists because a single fraudulent or hacked network device can compromise overall network security [49] and the amount of interconnectivity with ZigBee and Z-Wave raise concerns given their inherent security risks [37]. Thus, vetting communication device identity is critical to overall security. Regular operations of a typical communication network experiences many devices requesting network access. Passwords and keys required to gain access can be shared or forged, however the physical properties of a given device are inherently harder to forge.

Reliable network security involves considering multiple layers of access and interfacing between components and users. Devices, their operations, and applications for networks can be characterized by the seven layer Open System Interconnection (OSI) model, Table I-1. As one progresses from the Physical (PHY) layer to the Application layer, an increasing number of trust assumptions are made [50]. Historically, security has not adequately considered the physical attributes of devices themselves. Rather, much emphasis and research on network security and unauthorized access detection occurs at the Application, Network and Data Link layers [51–60], and Application Layer [61].

Table I-1: OSI Model, adapted from [62–65].

	DATA	LAYER	DESCRIPTION	EXAMPLE
HOST LAYERS	Data	Application	Process to access network	End User
		Presentation	Formats data for application layer, and encrypts data	Syntax, data manipulation
		Session	Interhost connections, session establishment	Synching
	Segments	Transport	End-to-end connections	TCP, host-to-host
MEDIA LAYERS	Packets	Network	Controls subnet, decides physical path for data, IP	Packets, routing
	Frames	Data Link	Transfer of data between nodes over physical devices	Frames, MAC addresses
	Bits	Physical	Transmission and reception of media, signal; physical devices.	Cables, devices, physical mediums, transmission methods

PHY features are considered as an additional level of security for more robust security systems and rogue device authentication [19]. For improved security and monitoring of device operations, it is desirable to collect and monitor identifiable features possessing qualities of *universality*, *distinctiveness*, *permanence*, and *collectability* [18, 66]. Moreover, these feature qualities are akin to biometric features [67–70]. AFIT’s RF-DNA Fingerprinting is one proven method for exploiting biometric-like features of electronic devices and was therefore of interest for this research.

1.2 Radio Frequency Fingerprinting

Broadly, there are two PHY-layer based security approaches that have been applied: 1) the addition of physically traceable objects to devices [71–73], and 2) the exploitation of inherent and unique features in device signals through RF

Fingerprinting [18, 74–76]. A variety of research has been conducted in the area of RF Fingerprinting – c.f. [49, 51, 66, 75–118], but each generally follows a similar procedure whereby fingerprint features are extracted from device emissions. In general, RF Fingerprinting processes involve 1) selecting Regions of Interest (ROIs) within a given signal response, 2) computing features from each ROI, 3) computing fingerprints from each feature, and 4) training classifier models to discriminant on these features [102]. RF Fingerprinting research has considered various wireless communication devices, including IEEE 802.11 (Wi-Fi) [92, 96, 97, 106, 119, 120], IEEE 802.16 (WiMAX) [98], 802.15.4 (ZigBee) [49, 89, 91, 113, 121, 122], Z-Wave [49, 123], Satellite Communication (SatCom) [124], Global System for Mobile Communications (GSM) cellular phones [101, 125], IEEE 802.15 Bluetooth [86], Ethernet [77, 126, 127], and Radio Frequency Identification (RFID) [78, 109].

Of specific interest in this research was the RF-DNA Fingerprinting method as codified by Cobb et al. [18, 19] and extended by work in [74]. As adopted here, the RF-DNA Fingerprinting process considered statistical features computed in each ROI of the instantaneous amplitude, frequency and phase responses [18]. RF-DNA has been employed in many applications [18, 19, 49, 74, 89–93, 97–99, 101, 113, 121, 128] and shown efficacy for both cross-model (different manufacturers) [101] and like-model (same manufacturer, same model, different serial number) device discrimination [92].

RF-DNA Fingerprinting embodies Wittgenstein’s [129] proposition that “in order to know an object, I must know not its external but all its internal qualities,” by augmenting the current external security measures via characterizing the internal

qualities. However, it should be stated that any measurements are model-based observations of the real phenomena [130], or as Heisenberg stated [131], "We have to remember that what we observe is not nature herself, but nature exposed to our method of questioning." Thus, RF-DNA Fingerprinting provides a reflection of the operating condition of RF devices, which has been further explored by directly analyzing integrated circuits (ICs) in [104].

1.3 Technical Motivation

RF Fingerprinting research has primarily focused on applications [49, 74, 78, 86, 89, 91, 92, 96–98, 106, 109, 113, 119, 121, 125] with classifier model development [19, 51, 91, 92, 132] and Dimensional Reduction Analysis (DRA) [49, 89, 113, 132] as secondary objectives. AFIT's RF-DNA work has previously considered four classification methods: Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) [90], Generalized Relevance Learning Vector Quantized-Improved (GRLVQI) [51], Learning from Signals (LFS) [133], and Decision Trees/Random Forests [134]. Additionally, since RF-DNA generally considers many fingerprint features, e.g. $N_F = 729$ features for the ZigBee dataset of [91], DRA has been of interest to select relevant subsets of features.

Various unresolved issues exist in RF Fingerprinting research and herein extensions are made to the RF-DNA process itself, classifier development, and DRA methods. Three previously unresolved issues related to DRA for RF Fingerprinting are addressed in Chapter IV: 1) understanding the appropriate use of p-values and test

statistic values when using distribution based DRA methods [49]; 2) developing MDA classifier-based DRA methods [135], which were previously dismissed [51, 89, 91, 92, 113, 134]; and 3) the development of quantitative dimensionality assessment methods to determine the number of features to consider [49, 135]. Recent RF-DNA efforts have considered a GRLVQI classifier, e.g. [51, 92, 100]; Chapter V addresses three general issues in GRLVQI: 1) extending the algorithm to consider non-Euclidean distance measures; 2) determining optimal algorithm parameter settings; and 3) creating a generalizable derivative skeleton to support algorithm improvements. Although the RF-DNA process is mature and proven, slight improvements to its operation are proposed in Chapter VI by leveraging techniques in Simulation research [136]; therefore, an autocorrelation based automation approach for selecting the number of ROI sub-regions is introduced.

1.4 Research Contribution

Table I-2 provides a summary and mapping of the contributions in this research, “Current Research,” to previous related research, “Prior Work.” In Table I-2, the × symbol indicates that a technical area was addressed.

Table I-2: Relational mapping between technical contributions in previous related work and *current* research contributions. The × symbol denotes areas addressed.

Technical Area	Prior Work		Current Research	
	Addressed	Ref #	Addressed	Ref #
ZigBee	×	[89, 91, 113, 121, 122]	×	[49, 135, 137]
Z-Wave			×	[49, 137]
Classification/Verification Processes				
MDA/ML	×	[18, 19, 89–91, 97, 101, 105, 113]	×	[49, 135]
GRLVQI	×	[51, 92, 97, 100, 128]	×	[49, 137]
LFS	×	[88, 92, 93, 94, 119, 133]		
Random Forests	×	[126]		
Dimensionality Reduction Analysis (DRA)				
MDA/ML	×	[18, 19, 51, 89–92, 113, 121]	×	[49, 135]
GRLVQI	×	[51, 92, 100]	×	[49, 135, 137]
LFS	×	[88, 92, 133]		
Random Forests	×	[132]		
KS-Test	×	[89, 91, 113, 121]	×	[49, 135]
F-Test			×	[49, 135]
Qualitative Dimensionality Assessment	×	[89, 91, 113, 121, 132]	×	[49, 135]
Quantitative Dimensionality Assessment			×	[49, 135]

1.5 Document Organization

This dissertation is subsequently organized as follows: Chapter II presents background literature on PHY layer device identification, RF signals, RF-DNA, the ZigBee devices under analysis, data collection, and pattern recognition. Chapter III presents the baseline classifier methods used in this study: MDA and GRLVQI. Chapter

IV reviews and develops DRA methods for application to RF-DNA. Chapter V presents improvements and modifications to GRLVQI, including a derivative framework to incorporate non-Euclidean distance measures and an optimization to method to determine algorithm parameter settings. Chapter VI presents concepts from simulation studies research and considers extensions to the RF-DNA process. Chapter VII concludes the dissertation. Appendices A through M, which provide additional results supporting concepts and conclusions in this dissertation, are provided following Chapter VII.

II. Background

Research has been proceeding to develop a line of...products that establishes new standards for quality, technological leadership, and operating excellence.

–MICHAEL KRAFT

This chapter provides the foundation for understanding physical (PHY) layer security of communication devices, Radio Frequency Distinct Native Attribute (RF-DNA) Fingerprinting, ZigBee and Z-Wave signals under analysis, and particulars of signal collection and RF-DNA feature extraction.

2.1 Introduction

This chapter is organized as follows. First, a general discussion on wireless networks and a specific discussion on ZigBee and Z-Wave devices are presented in Section 2.2. Then a discussion on PHY security and device identification is presented in Section 2.3. Finally, the RF-DNA Fingerprinting process is presented and discussed in Section 2.4.

2.2 Signals of Interest: Wireless Networks

Figure II-1 presents a conceptualization of basic digital communication occurring between two devices [64, 138]. In operation, a software application initiates the communication of a data packet, as the packet proceeds through each layer of the Open Systems Interconnection (OSI) model more information in the form of headers, addresses and etc., are added at each layer regarding the device properties, bit-level identity, communication properties, data handling information, and etc. [138]. After passing

through the OSI layers, the digitally formatted signal is transmitted over some medium (wired or wireless) and received by another device. The receiving device collects the signal and reverses the digital formatting process, including the removal of headers at each layer to determine how to handle the received data [138].

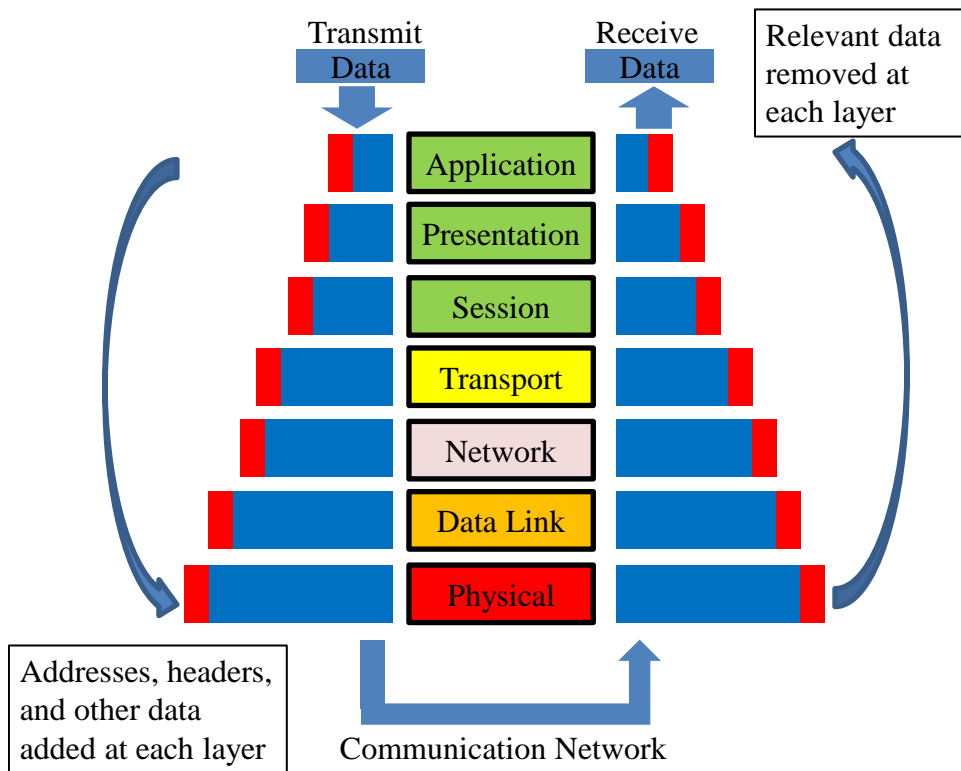


Figure II-1: General operations of digital communication, adapted from [64, 138].

Various technical standards exist that govern the operation of a wide variety of communication networks. Of interest herein are ZigBee wireless personal area networks (WPAN) which are governed by the WPAN working group (IEEE 802.15); one of 25 IEEE 802 standard subgroups for area networks [139]. The IEEE 802.15 working group also includes Bluetooth (IEEE 802.15.1), coexistence (IEEE 802.15.2), high rate WPANs (IEEE 802.15.3), the low rate WPANs (IEEE 802.15.4), mesh networking (IEEE

802.15.5), body area networks (IEEE 802.15.6), and visible light communication (IEEE 802.15.7) [139, 140]. Due to their operating characteristics, ZigBee devices fall under the IEEE 802.15.4 subgroup.

2.2.1 IEEE 802.15.4 ZigBee Devices

ZigBee devices are low-cost, low-data rate, low-complexity wireless communication devices which can function at nominal ranges of 10-100 meters and support networks containing up to 65,000 devices [38, 39, 141]. Given these attributes, ZigBee devices are employed for various tasks and are consequently connected to more devices in the physical world than any other wireless technologies [37, 38]. Various ZigBee applications include maritime environments [142], smart thermostats [37], electronic door locks (e.g. Kwikset SmartCode) [37] and security devices [143], smartphone controlled doorbells [144, 145], building automation and control [37, 46, 146], greenhouse monitoring [147, 148], healthcare [149, 150], energy management [151–153], HVAC (heating, ventilation, and air conditioning) operations [143], smart metering [154–156], electricity theft detection [48, 157], smart homes and smart appliances [158, 159], waste-water management [160], chemical plant automation [161], electric substation automation [162], and meter reading [163]. Many of these applications are in areas considered ‘critical infrastructure (CI),’ the interruption of which can threaten lives, disable governments, affect economies, and damage ecological systems [3]. Due to the functional interconnectedness of such complex systems, a compromise at one point can compromise the overall system security [6].

ZigBee network security frequently incorporates a 128-bit advanced encryption standard (AES), 16-bit cyclic redundancy check (CRC) for data protection, and cipher block chaining message authentication code (CBC-MAC) for authentication [38]. However, despite their near ubiquity and security precautions, ZigBee networks are vulnerable to intrusion through readily available ‘hacking tools’ such as KillerBee [37] or Packet-in-Packet approaches [164]. Unfortunately, current ZigBee security mechanisms frequently neglect the PHY layer where much malicious activity occurs [51]. PHY layer protection involves device identification and authentication; various reasons exist for examining this layer, including access control, augmenting other security measures, authentication, intrusion detection, malfunction detection, and rogue access, among other applications [66, 165, 166].

When considering ZigBee devices as an RF-DNA problem, knowledge of the underlying standard, IEEE 802.15.4 [121, 167], is important in order to determine how and with what signal to create RF-DNA fingerprints. IEEE 802.15.4 has defined PHY, Media Access Control (MAC), and Network (NWK) layer specifications. In the operation of transmitting a burst signal, a ZigBee device transmission at the PHY layer involves a structure, termed a PHY Protocol Data Unit (PPDU); the PPDU contains a defined Synchronization Header Response (SHR), a 8-bit PHY Header Response (PHR), in addition to a variable length ‘payload’ contained in the PHY Service Data Unit (PSDU) which consists of a MAC sublayer frame [91]. The underlying ZigBee PHY layer packet structure is conceptualized in Figure II-2.

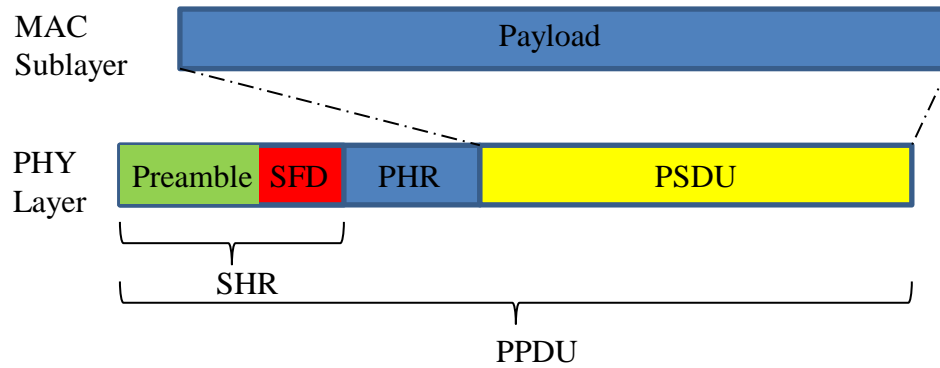


Figure II-2: ZigBee PHY layer packet structure, adapted from [167].

Different ZigBee device formats exist and the SHR varies in length and duration for different ZigBee PHY options, i.e. frequency (868MHz to 2.4GHz), and shift keying approach [168]. ZigBee devices can employ amplitude shift keying (ASK), binary phase shift keying (BPSK), or quadrature phase shift keying (QPSK), as seen in Table 3.4 of [168]. However, while the format of each region changes per keying method, the use of each region is consistent across ZigBee devices: the preamble is used for synchronization between devices, and the SFD region used to indicate the end of the SHR and the start of the PHR [168].

Of specific interest herein are Texas Instruments CC2420 2.4GHz ZigBee devices which employ QPSK, [91]. These devices have a defined 128 μ s duration preamble of 4 octets (4-bytes) which contain 8 zeros each, and a 1 octet (1-byte) defined SFD containing 2 hexadecimal symbols [168]. The ZigBee SHR region format is presented in Table II-1. Four synchronization words (SWs) are defined as the last octet of the preamble and the SFD [167]; alternately, Farahani [168] lists possible SFD values of E5, or 11100101. The PHR region contains frame length information and is one 1 byte in length and ranges from 0 to 127 [168].

Table II-1: Zigbee SHR Region Format, adapted from [91, 167, 168].

REGION	SHR									
	PREAMBLE								SFD	
HEXA-DECIMAL VALUE	0	0	0	0	0	0	0	0	7	A
BINARY	0000	0000	0000	0000	0000	0000	0000	0000	0111	1010
CC2420	Zeros						SW0	SW1	SW2	SW3

2.2.1.1 ZigBee Data Collection Experiment

The ZigBee dataset under analysis is a four class authorized device classification model development problem with six additional rogue devices for verification [91]. Signals from the ZigBee devices were collected in three different environments: ‘CAGE,’ signals in a Ramsey STE3000B RF shielded anechoic chamber; ‘LOS,’ line of sight signals in an office hallway, denoted by A in Figure II-3; and ‘WALL,’ signals collected behind a wall, denoted by B in Figure II-3 [91].

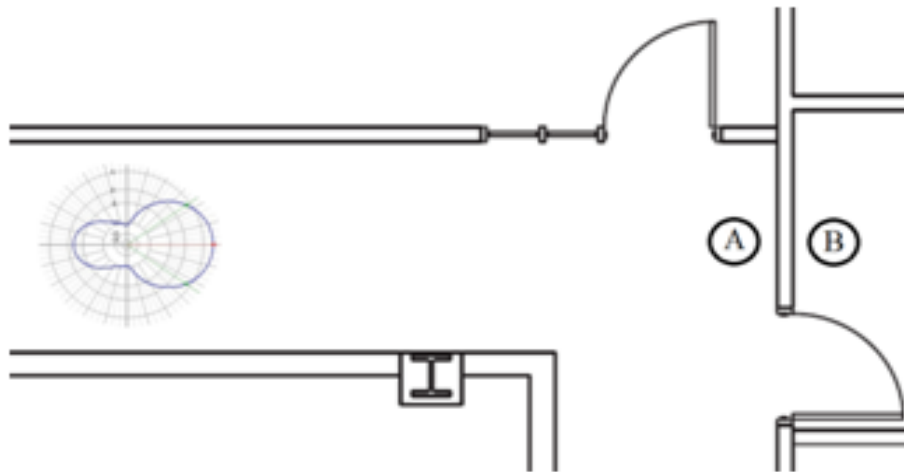


Figure II-3: Conceptualization of ZigBee data collection, from [91].

Table II-2 describes the data collection experiment and the data available for each ZigBee device. The four devices used for model development (Dev1 – Dev4) had data collected in all three environments [91]. However, data from the rogue devices (Dev5 – Dev10) was only collected in one or two environments [91]. For operation and ensuring that the number of observations by rogue device is consistent, the WALL collections of devices 5-7 are considered as additional devices [91].

Table II-2: ZigBee Collected Data, adapted from [91].

	ZigBee ID	CAGE	LOS	WALL
AUTHORIZED	Dev1	X	X	X
	Dev2	X	X	X
	Dev3	X	X	X
	Dev4	X	X	X
ROGUE	Dev5		X	X
	Dev6		X	X
	Dev7		X	X
	Dev8	X		
	Dev9	X		
	Dev10	X		

ZigBee burst signal data was collected by Dubendorfer [91] using an Agilent Receiver to collect burst transmission from the ten Texas Instruments CC2420 2.4GHz ZigBee devices. The ZigBee devices were setup to transmit at 2.4GHz, within the Agilent receiver's 20.0MHz to 6.0GHz range and 36.0MHz bandwidth [91]. For each device, 1000 burst responses of the SHR and PHR regions were collected under three different operating conditions [91].

2.2.2 IEEE 802.15.4 Z-Wave Devices

While the ZigBee device dataset is representative of many applications, it only considers one type of device. Therefore, consistent with [49], in addition to the ZigBee devices Z-wave devices are considered as an extension to this research. Both ZigBee and Z-Wave devices are small, low-cost wireless communications devices, however differences exist between ZigBee and Z-Wave in, primarily, standards and security [169].

While ZigBee devices employ an IEEE standard for industrial, residential and sensor monitoring and automation, Z-wave devices employ proprietary standard developed by ZenSys for, primarily, residential automation [170–172]. While ZigBee and Z-Wave are similar in concept and possible use, differences exist in security, operating frequency, data rate, and latency as seen in Table II-3. Primarily, Z-Wave is considered less secure than ZigBee due to Z-Wave originally lacking built in encryption [170]. Additionally, the Z-Wave standard is proprietary and not publically available, unlike ZigBee [172].

Table II-3: ZigBee versus Z-Wave, adapted from [170, 172].

	Z-Wave	ZigBee
FREQUENCY	906 MHz	2.4 GHz
BIT RATE	40 Kbits/s	250 Kbits/s
SECURITY	None (200 and 300 series models) AES 128 (400 series models)	IEEE 802.15.4 security standards
LATENCY	~1000 ms	50-100 ms
RANGE	30-100 m	10-100 m
MESSAGE SIZE (BYTES)	64 (max)	127 (max)

Z-Wave follows a similar ISO architecture to ZigBee, and similarly has a predefined preamble and SoF [173]. A conceptualization of the Z-Wave PHY packet structure is presented in Figure II-4, for RF-DNA the preamble is again considered as the ROI in the signal. Z-Wave also includes a payload-based home identification (32-bits) and source identification (8-bits) [172].

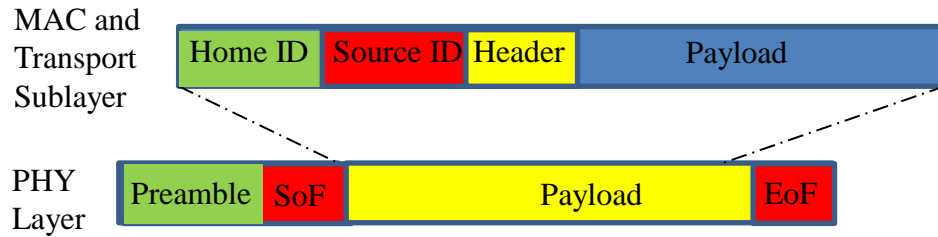


Figure II-4: Z-Wave PHY layer packet structure, adapted from [173].

For purposes herein, three Aeotec Z-Stick S2 transmitters, consistent with [174], were employed as described by [49, 123]. A total of 230 Z-Wave bursts were collected at 2 Msps, with the preambles being the first 8.3 ms of each burst. Z-Wave data was collected under LOS conditions with the Z-Wave devices placed 10 cm from a vertically-oriented LP0410 log-periodic antenna, which was connected via a Gigabit Ethernet cable to an USRP-2921 RF input [49]. Amplitude-based leading edge detection was employed with a -6 dB detection threshold to detect and extract the bursts from the background noise [49]. The collected signal had a Signal-to-Noise Ratio of $SNR = 24$ dB and was like-filtered [49].

2.2.3 Post-Collection Data Manipulation

After collecting the ZigBee RF emissions, Dubendorfer [91] converted the files to MATLAB format. Since the SHR and PHR regions begins each ZigBee transmission, and are not changed between devices, the RF-DNA process was applied to this region of the ZigBee transmission [91]. First, Dubendorfer [91] detected the bursts from the ZigBee devices, which comprise the signals of interest. After digital filtering through a Butterworth baseband filter, additive white Gaussian noise (AWGN) was included to create a range of operating points (16) between $SNR = 0$ and $SNR = 30$ dB using five independent noise realizations per device [91]. A similar approach was considered for the Z-Wave devices, where AWGN was added to collected signals to achieve desired operating points of $SNR \in [0\ 24]$ dB in 2 dB steps [49].

2.3 Physical Layer Device Identification

Because PHY layer characteristics are associated with the physical properties of devices, they are naturally harder to spoof than characteristics associated with other OSI levels [175]. PHY layer security consists of two broad approaches for exploiting RF-emission features: 1) adding a physical object to an electronic device, such as an RF-Certificate of Authority (COA), or 2) exploiting inherent emission features of electronic devices, such as RF-DNA. A brief review of the various approaches is considered to illustrate the benefits of the RF-DNA approach.

2.3.1 RF Device Emissions

Both intended and unintended emissions occur across the electromagnetic spectrum in a variety of forms; intentional emissions can range from light emitted from a light bulb to wireless communications. Unintended emissions are also emitted from a variety of sources; one commonly experienced form of unintended electromagnetic emissions occurs through light pollution which makes viewing the night sky difficult in urban areas [176]. Since the 1970s man-made noise from unintended emissions has increased due to the proliferation of electronic devices [177]. Electronic device emissions have security [178], safety [179], interference and communications [180] ramifications. Although shielding and design are used to reduce unintended emissions, the underlying physics of electronic devices precludes their elimination [180, 181].

RF emissions can emanate from both intended and unintended radiators [182]; unintended RF emissions emanate from normal operations and are caused by transistor switching, current flow, integrated circuit (IC) activity, in addition of other electromagnetic effects [19, 183]. Although unintended RF emissions are a generally considered a source of interference, they are also useful for device identification between disparate devices [184]. When devices from the same production run are considered, production-induced variations result in devices being within production tolerances yet having different RF emissions [19]. Although exploiting intentional device emissions is of concern herein, exploring methods used to exploit both unintended and intended emissions adds important background knowledge for this research.

Four leading RF-based device identification methods have been proposed: Radio Frequency Identification (RFID), Physical Unclonable Functions (PUFs), RF Certificates of Authenticity (RF-COA), and RF Fingerprinting. Of these, only RF Fingerprinting exploits signals that inherently emanate from the device, while the other three methods requiring the addition of components to the underlying devices.

2.3.1.1 Radio Frequency Identification (RFID)

RFID is a tracking technology seen in some RF physical layer security schemes. RFID involves placing a ‘tag’ on a device for tracking; each tag is an identifier antenna circuit based on RF communication between the antenna and a scanner [185, 186]. RFID antennae can be either powered and actively emitting or unpowered and emitting only when scanned [71]. RFID has seen applications in many commercial and warehouse applications where products and parts are tracked [186]. RFID does have known issues, including: interference [187], and obviously the practical issue of requiring an RFID antenna to be knowingly placed (visible or otherwise) on an object in order for it to be scanned.

2.3.1.2 Physical Unclonable Functions (PUFs)

PUFs offer two techniques for authentication: 1) augmenting an IC with internal measurement circuitry, and 2) adding a grid of capacitive sensors onto the top IC layer [19]. Both of these PUF approaches require physical IC manipulation and therefore are prohibitive to exploring due to legacy ICs being in operational use.

2.3.1.3 RF Certificates of Authenticity (RF-COA)

RF-COAs are another attempt to add identifying characteristics to electronic devices. RF-COAs extend the RFID concept by placing small, unique, three-dimension antennae comprising of randomly shaped conductors and dielectric components, COAs, onto electronic device to create a unique identifiable RF signal [73]. The philosophy of this approach is that where unique COAs would be issued by manufacturers of objects and software to confirm their provenance [73]. In essence, RF-COAs are a combination of PUFs and RFID, where the RF-COAs are read by an external RFID type of reader [19]. The obvious impediment is the emplacement of the RF-COAs on devices already in operation, the additional cost of extra components, and additional considerations in the design and fabrication process. The ease of spoofing is also a known issue with the COA approach [73].

2.3.1.4 RF Fingerprinting

RF Fingerprinting refers to one of two processes: characterizing the RF environment devices operate in, c.f. [188, 189], or identifying devices based on differences in transmitted signals resulting from differing characteristics, due to production and life style variations, among various devices [79]. Of interest herein is that AFIT RF-DNA RF Fingerprinting process which is unique in RF Fingerprinting in that it applies statistical methods of feature extraction and classification to the RF Fingerprinting process [133]. RF-DNA has been explored for both inter-device variations, e.g. differentiating similar devices from different manufacturers [190], and

intra-device variations, e.g. differentiating devices as the serial number level [91, 190]. In operation, the AFIT RF-DNA process consists of two parts, the signal collection aspect (which involves various signal collection equipment) and the processing aspect (which occurs within MATLAB) [190], Figure II-5.

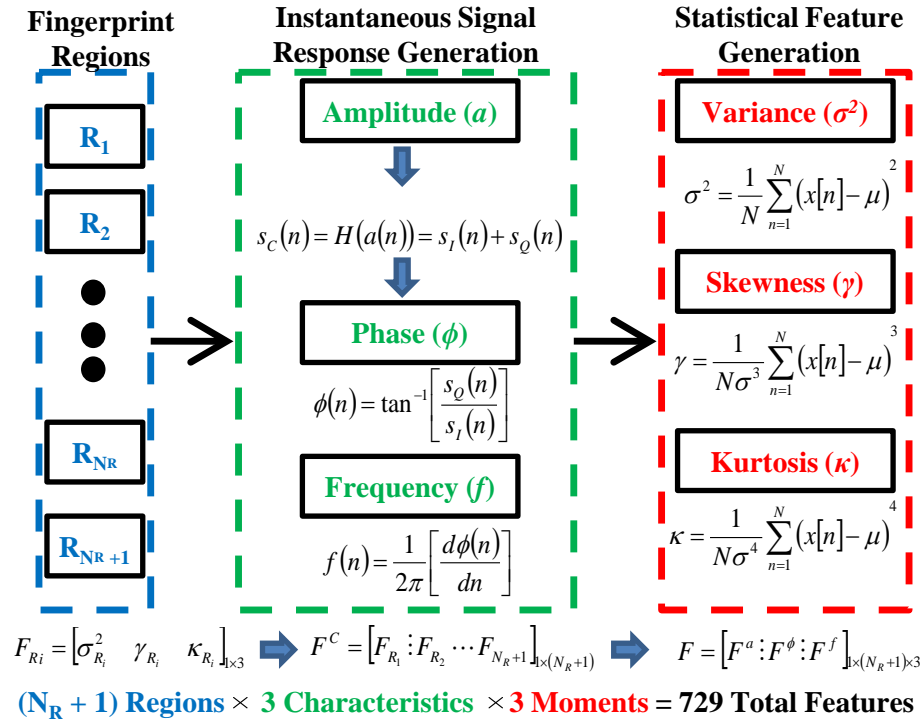


Figure II-5: RF-DNA Fingerprinting Architecture, adapted from Cobb et al. [19].

After collection, the data is digitally filtered and manipulated to create samples at various SNR levels. Following this, RF-DNA fingerprints are computed and various classification schemes are applied for model development and verification of the models is explored using rogue devices. RF-DNA involves extracting fingerprints from RF emissions; in a manner, akin to biometrics in finding unique attributes of electronic

devices. A visualization of computing RF-DNA fingerprints from sampled-time ZigBee SHR data is presented in Figure II-6.

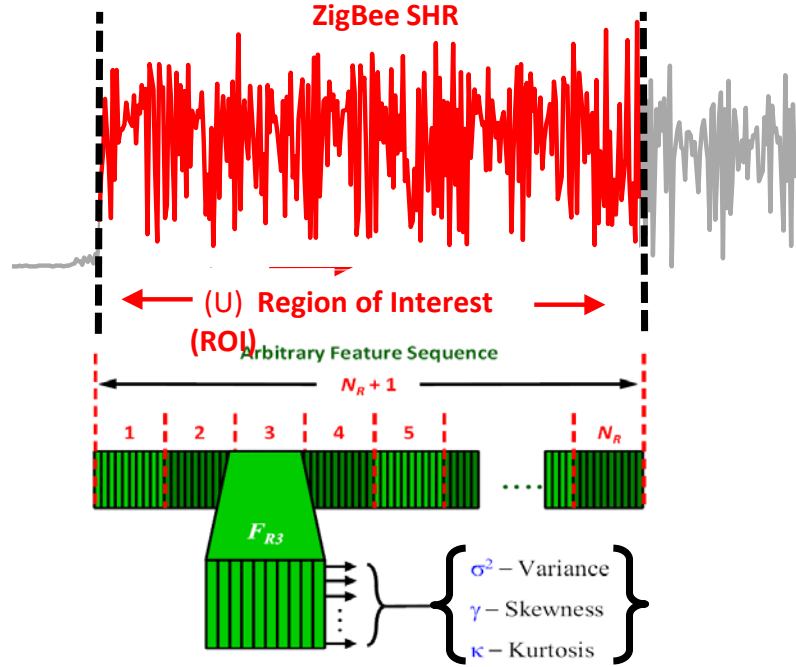


Figure II-6: Traditional RF-DNA Feature Extraction Approach as Applied to ZigBee Devices, adapted from [91].

2.4 1D Time Domain (TD) RF-DNA Fingerprints

After dividing the collected and processed data's ROI into bins, the signal's instantaneous amplitude (a), phase (ϕ), and frequency (f) response are computed for each [89, 91, 128]. When considering the region of interest (ROI) of the sampled signal as a complex I-Q equation,

$$s[n] = s_I[n] + js_Q[n], \quad (2.1)$$

the RF-DNA fingerprint elements can be computed thusly [91]:

$$a[n] = \sqrt{s_I^2[n] + s_Q^2[n]}, \quad (2.2)$$

$$\phi[n] = \tan^{-1} \left(\frac{s_Q[n]}{s_I[n]} \right), \text{ for } s_I[n] \neq 0, \quad (2.3)$$

$$f[n] = \frac{1}{2\pi} \left(\frac{d\phi[n]}{dt} \right), \quad (2.4)$$

consistent with general formulations found in [64, 191, 192]. Per Dubendorfer [91], (2.2)–(2.4) are normalized through subtracting the mean and dividing by the maximum,

$$\bar{g}_c[n] = \frac{g[n] - \mu_g}{\max(g_c[n])}, \quad (2.5)$$

where g in (2.5) represents the respective RF-DNA fingerprint elements in (2.2)–(2.4) for $n = 1, 2, \dots, N_S$, where N_S represents the number of samples in the region, and μ_g represents the mean of the g -th fingerprint element.

RF-DNA fingerprints features are then extracted from the normalized amplitude frequency and phase. The considered RF-DNA features are 2nd, 3rd, and 4th mathematical moments of variance (σ^2), skewness (γ), and kurtosis (κ) [90, 91]. Standard deviation can also be computed as an RF-DNA fingerprint, and was applied by [51]; however, as it is necessarily highly correlated with variance, it was not applied to ZigBee signals by Dubendorfer [91], and it will not be examined herein.

Considering the 2nd to 4th mathematical moments enables an understanding of distributional properties within each bin, respectively the variability about the mean (variance), asymmetry about the mean (skewness), and distribution curvature (kurtosis), [193–195]. Mathematical moments have also seen similar applications are seen in other

domains, cf. [196–201]. Computed, skewness values are centered at 0 which indicates no skewness about the mean; skewness values are then either positive, for a left sided distribution, or negative, for a right-sided distribution [202]. Kurtosis values indicate pointedness or flatness of a distribution with values of either $\kappa = 3$, termed mesokurtic, $\kappa < 3$, termed platykurtic (flatter), and $\kappa > 3$, termed leptokurtic (more pointed) [202]. Consistent with RF-DNA features of σ^2 , γ , and κ are computed for N total samples through the following formulas:

$$\sigma^2 = \frac{1}{N} \sum_{n=1}^N (x[n] - \mu)^2, \quad (2.6)$$

$$\gamma = \frac{1}{N\sigma^3} \sum_{n=1}^N (x[n] - \mu)^3, \quad (2.7)$$

$$\kappa = \frac{1}{N\sigma^4} \sum_{n=1}^N (x[n] - \mu)^4, \quad (2.8)$$

where,

$$\mu = \frac{1}{N} \sum_{n=1}^N x[n], \quad (2.9)$$

and $x[n]$ represents an n^{th} feature vector element from the amplitude, phase, or frequency response [91].

Combined together, the RF-DNA features are arranged in a vector as

$$F_{R_i} = [\sigma_{R_i}^2 \quad \gamma_{R_i} \quad \kappa_{R_i}]_{1 \times 3}, \quad (2.10)$$

for each observation $i=1,2,\dots, N_R+1$, where N_R refers to the total number of observed sequences with the additional observation refers to statistics computed over the entire

signal characteristic. When considering an entire characteristic's features, (2.10) extends to

$$\mathbf{F}^c = \begin{bmatrix} F_{R_1} \\ F_{R_2} \\ \vdots \\ F_{R_{N_R}} \\ F_{R_{N_R+1}} \end{bmatrix}. \quad (2.11)$$

When considering the amplitude, frequency, and phase fingerprints, (2.10) and (2.11) are extended through concatenations:

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}^a \\ \mathbf{F}^\phi \\ \mathbf{F}^f \end{bmatrix}. \quad (2.12)$$

2.4.1 ZigBee and Z-Wave RF-DNA Fingerprinting

For all ZigBee devices of interest, authorized or rogue, $N_F = 729$ total features were computed from the collected time domain burst signal [91]. This corresponds to 3 statistical features and 81 bins (78 separate regions, and 3 averaged regions for the entire signal). For each feature, 1000 exemplars were computed each for CAGE, LOS, and WALL [91]. Additionally, data was available for 16 SNR levels, $SNR \in [0 \ 30]$ dB, with each having five different noise realizations.

For classifier model development training and testing, the dataset of authorized device is separated into upper and lower halves; these were 'interleaved' meaning every odd-indexed point was selected for training and every even-indexed point was selected for testing. In this form, the training and test sets for ZigBee devices both consisted of 500 CAGE observations, 500 LOS points, and 500 WALL points.

In operation, these structures are organized as a four dimensional data structure with N_{FP} represents fingerprint observations; N_{Feats} , features; N_{Nz} , noise realizations; and N_C classes. For ZigBee data, the structure is of size $3000 \times 729 \times 5 \times 4$. For interpretation, not everyone has mental familiarity with four dimensional structure, an example of what this means would be: there are 3000 points associated with feature 1 of noise realization 1 of device 1 and so on. For the rogue devices, 1000 samples were collected in the respective environment; for data storage and dimensionality concerns, this is considered as 3000 points with only the first 1000 correspond to fingerprint data, and the remaining 2000 being zeros.

For the Z-wave devices under consideration, 230 LOS observations were collected and a total of 189 RF-DNA features were computed for $N_{FP} = 230$, $N_{Feats} = 189$, $N_{Nz} = 2$, $N_C = 3$; thus, the Z-Wave data structure is of size $230 \times 189 \times 2 \times 3$. While the ZigBee dataset is of primary interest herein, the Z-Wave dataset will permit quick algorithmic development due to its smaller size. Additionally, the Z-Wave dataset will allow generalization of results to more than one signal of interest.

III. Statistical Pattern Recognition

Can the truth be learned? With this question we shall begin.

—SØREN KIERKEGAARD, 1813-1855

The nature of the physical world and how objects are differentiated and created has concerned man since time immemorial: e.g. pre-Socratic *physiologoi* such as Anaxagoras, Anaximander, and Democritus thought on the origin and nature of phenomena [203, pp. 14–28; 203, pp. 249–267; 204, pp. 82–86; 205, pp. 350–359]. Systematic methods of pattern recognition begin with Aristotelian thought, with Aristotelian metaphysics concerned with the nature of being [203, p. 139], Aristotelian category theory [206], and questions of classification in Eastern thought, e.g. verse 2 and 6 of the *Tao Te Ching* and verse 61 of the *Hua Hu Ching* [207, 208]. Locke considered thinking as part sensation and part reflection, extending Descartes’ duality of mind with the observation that the mind considers either “sensations” or “reflections” [209, 210], similarly Hume viewed that one needs to experience something before one can visualize that something [211]; in essence these propositions echo training and testing problems in pattern recognition. Pattern recognition is critical to both every day and computational tasks [212], and broadly covers classification of objects, clustering, and recognizing variables and patterns of variables [213]. The term statistics has also become associated with data analysis. Originally referring to a science of politics [214], and descended from the Latin *statista*, meaning “political state” [215], its meaning has shifted to become synonymous with data analysis and distributional measures [215].

3.1 Introduction

This chapter is organized as follows, Section 3.2 discusses Multiple Discriminant Analysis (MDA), Section 3.3 discusses the Learning Vector Quantization (LVQ)–family of algorithms, including Generalized Relevance Learning Vector Quantization Improved (GRLVQI), and Section 3.4 discusses performance assessment methods of interest to RF INT. Of particular interest herein are statistical methods applied to pattern recognition tasks, especially those used for supervised clustering or ‘classification’ where patterns are compared with a set of known classes [213]. This differs from unsupervised classification, commonly known as ‘clustering,’ where known predefined groups do not exist [213]. Additionally, supervised classification for RF Distinct Native Attribute (RF-DNA) problems considers two parts: classification and verification [19]. The first part of classification involves the classifier model development stage where the primary concern is a “one vs many” problem of known group identities with the goal to create a classifier model that effectively discriminates between authorized devices [19]. Verification involves vetting the classification model by how well they recognize authorized and non-authorized devices (*rogue*), in a “one versus one” claimed identity problem [19].

Various classification methods exist; herein we are primarily concerned with methods previously employed for RF-DNA features, namely MDA and the GRLVQI algorithm. Both MDA and the LVQ-family of algorithms are described below; MDA is a linear method whereas LVQ methods are nonlinear approaches that incorporate various nearest neighbors, neural network and nonlinear concepts.

Figure III-1 presents a conceptualization of differences in classifier paradigms between MDA and LVQ approaches, showing MDA minimizing inter-class differences while maximizing intra-class difference and LVQ minimizing inter-class prototype vector magnitudes and maximizing the distance between intra-class prototype vectors. In describing both MDA and the various LVQ methods, the following general notion will be used: the input data matrix is defined as \mathbf{X} which has N_{tot} total observations (rows) and N_F data features (columns). This will additionally be considered for N_C classes.

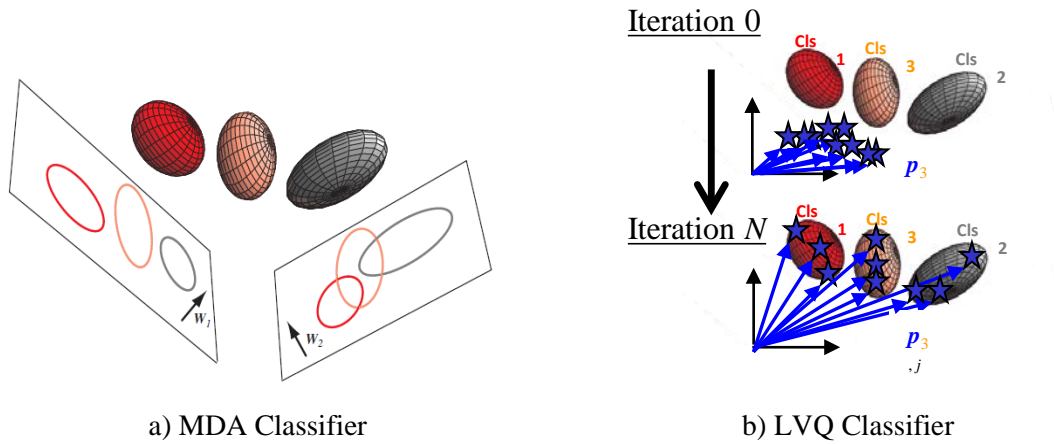


Figure III-1: Conceptualization of a) MDA class projections from [216] and b) LVQ prototype development as adapted from [51, 216].

3.2 Multiple Discriminant Analysis

MDA extends Fisher's linear Discriminant Analysis (DA) to multiple classes [216, pp. 121-124]. DA and MDA are frequently used for predictive/classification and descriptive/clustering tasks and are frequently applied to tasks and domains ranging from ecology [217, 218], civet coffee authentication [219], behavioral sciences [220], marine

data analysis [221, 222], muzzle flash identification [223], and MDA/Maximum Likelihood (MDA/ML) methods for RFINT [88, 92–94, 119, 133, 224]. MDA and DA also frequently compare favorably (in either/or accuracy and computation time) to more complicated statistical methods, such as neural networks, logistic regression, support vector machines, naïve Bayes classifiers and LVQ approaches, c.f. [51, 92, 225–229]. Current research extensions and variants of DA and MDA also exist, these include extending MDA or DA to use other machine learning and statistical tools, such as kernels or nonparametric statistics [230–234].

MDA is a linear classifier based on Fisher’s 2 class method, but extended to multiple classes [235, 236]. Weight vectors are computed for sample based estimates using the Fisher criterion function for maximum discrimination,

$$\lambda = \frac{\mathbf{b}^T \mathbf{S}_b \mathbf{b}}{\mathbf{b}^T \mathbf{S}_w \mathbf{b}}, \quad (3.1)$$

which is a ratio of the between groups and within groups sum of squares with \mathbf{b} being the discriminant weights (eigenvectors) of $\mathbf{S}_w^{-1} \mathbf{S}_b$, and λ being the associated eigenvalue that equals the separation [237, 238]. To maximize λ with respect to \mathbf{b} , (3.1) can be treated as a maximization problem, $\max_b \mathbf{b}^T \mathbf{S}_b \mathbf{b}$ subject to $\mathbf{b}^T \mathbf{S}_w \mathbf{b} = 1$, by taking the partial derivative and setting equal to zero [239, 240]. Considering the Lagrangian,

$$L = \mathbf{b}^T \mathbf{S}_b \mathbf{b} - \lambda(\mathbf{b}^T \mathbf{S}_w \mathbf{b} - 1), \quad (3.2)$$

and taking the partial derivative of (3.2) with respect to \mathbf{b} ,

$$\frac{\partial}{\partial \mathbf{b}} (\mathbf{b}^T \mathbf{S}_b \mathbf{b} - \lambda (\mathbf{b}^T \mathbf{S}_W \mathbf{b} - 1)) = 2\mathbf{S}_b \mathbf{b} - 2\lambda \mathbf{S}_W \mathbf{b}, \quad (3.3)$$

one arrives at a problem similar to eigenvalues/eigenvectors [237, 238]. Setting (3.3) equal to zero yields,

$$(\mathbf{S}_b - \lambda \mathbf{S}_W) \mathbf{b} = (\mathbf{S}_W^{-1} \mathbf{S}_b - \lambda \mathbf{I}) \mathbf{b} = 0, \quad (3.4)$$

a common eigenvalue/eigenvector problem [216]. Taking the partial derivative of (3.2) with respect to λ gives,

$$\mathbf{b}^T \mathbf{S}_W \mathbf{b} = 1, \quad (3.5)$$

hence the eigenvector is scaled to unit variance.

The between class sum of squares S_b is defined as

$$\mathbf{S}_b = \mathbf{S}_T - \mathbf{S}_W, \quad (3.6)$$

with S_W , the within class scatter matrix, defined as

$$\mathbf{S}_{W_i} = \sum_{j=1}^{N_i} (\mathbf{X}_{ij} - \boldsymbol{\mu}_i)(\mathbf{X}_{ij} - \boldsymbol{\mu}_i)^T, \quad (3.7)$$

where $\boldsymbol{\mu}_i$ is the i^{th} group mean or centroids, and N_i are the total number of observations in the i^{th} group [237, p. 401]. The within groups sum of squares, assuming the covariance matrices of the classes are equal, is $S_\omega = S_{\omega_1} + S_{\omega_2} + \dots S_{\omega_c}$; and the total mean corrected sums of squares and cross products is defined as:

$$S_T = \sum_{i=1}^c \sum_{j=1}^{n_i} (\mathbf{X}_{ij} - \boldsymbol{\mu}_0)(\mathbf{X}_{ij} - \boldsymbol{\mu}_0)^T, \quad (3.8)$$

where $\boldsymbol{\mu}_0$ represents the grand mean vector [19, 216]. Data \mathbf{X} is then projected to an N_{df} dimensional discriminant space according to

$$\mathbf{G} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{N_{df}}]^T \mathbf{X}, \quad (3.9)$$

where

$$N_{df} = \min(N_C - 1, N_F), \quad (3.10)$$

which restricts the total number of discriminant functions [237, p. 401]. Although (3.10) is frequently specified as $N_C - 1$ [19, 51, 90, 91], such a reduction may not be appropriate if a small set of features is used or selected. The maximum number of discriminant functions to generate is determined by the eigenvalues of $\mathbf{S}_W^{-1}\mathbf{S}_b$. If the eigenvalues of $\mathbf{S}_W^{-1}\mathbf{S}_b$ are distinct, the number of linear composites will be bounded by rank of \mathbf{S}_b and, consequently, the rank of $\mathbf{S}_W^{-1}\mathbf{S}_b$ [237, p. 401]. Additionally, when the number of features exceeds the number of observations the covariance matrix is obviously singular, which can violate distributional assumptions and enable situations of complex discriminant loadings with further dubious underlying discriminant functions.

3.2.1 MDA Feature Relevance Ranking

Classifier-based feature relevance rankings from MDA are currently unexplored in RF-DNA methods with some research, e.g. [51, 91, 92, 113, 134, 241], even positing that one cannot extract feature relevance rankings from MDA. However, the method of discriminant loadings is one approach that directly computes the contributions of each data feature to the resultant discriminant functions.

Discriminant loadings reflect the contribution of each data feature to a given discriminant function and are analogous to principal component loadings [237, pp. 394-429]. Dillon and Goldstein [237] suggest that due to the unsuitability of the eigenvectors

to provide information of the contribution of each feature to the discriminant functions, one should therefore compute the loadings. It is of interest to examine the ‘contribution’ of each input feature to each discriminant function as means of screening data features. Occasionally, these values are reported in literature [242], but they are usually included to describe results. Dillon and Goldstein list discriminant loadings as the simple correlation between discriminant scores and the input data features [237, p. 414], and explicitly for the j^{th} discriminant function [237, p. 373]:

$$\mathbf{L}_j = \text{corr}(\mathbf{X}, \mathbf{b}_j \mathbf{X}) = \text{corr}(\mathbf{X}, \mathbf{X}) \mathbf{b}_j. \quad (3.11)$$

The statement of Dillon & Goldstein [237, p. 414], “...discriminant loadings for a variable...is the correlation between the function, \mathbf{G} from (3.9), and the variable...” and echoed in [237, pp. 372-373], is interpreted by [243] as:

$$\mathbf{L}_i = \text{corr}(\mathbf{X}, \mathbf{G}), \quad (3.12)$$

where we are really computing the correlation of \mathbf{X} with (3.9). Realizing that

$$\text{cov}(\mathbf{X}, \mathbf{b}^T \mathbf{X}) = \text{cov}(\mathbf{X}, \mathbf{X}) \mathbf{b}, \quad (3.13)$$

then the correlation expression in (3.12) can be rewritten as

$$\text{corr}(\mathbf{X}, \mathbf{b}^T \mathbf{X}) = \mathbf{D}_X^{-1/2} \text{cov}(\mathbf{X}, \mathbf{X}) \mathbf{b} \mathbf{D}_{\mathbf{b}^T \mathbf{X}}^{-1/2}. \quad (3.14)$$

where \mathbf{D}_X is a matrix of the diagonal entries of $\text{cov}(\mathbf{X}, \mathbf{X})$ and $\mathbf{D}_{\mathbf{b}^T \mathbf{X}}$ is a matrix of the diagonal entries of $\text{cov}(\mathbf{b}^T \mathbf{X}, \mathbf{b}^T \mathbf{X}) = \mathbf{b}^T \text{cov}(\mathbf{X}, \mathbf{X}) \mathbf{b}$ [243]. This further expands to

$$\text{corr}(\mathbf{X}, \mathbf{b}^T \mathbf{X}) = \text{corr}(\mathbf{X}, \mathbf{X}) \mathbf{D}_X^{1/2} \mathbf{b} [\mathbf{b}^T \text{cov}(\mathbf{X}, \mathbf{X}) \mathbf{b}]^{-1/2}. \quad (3.15)$$

One could feasibly scale MDA coefficients to ensure equal variance in all directions; therefore one area of related interest is how, if at all, MDA loadings are

possibly affected by scaling the projection matrix. Appendix A addresses this issue by presenting a lemma that proves MDA loadings are not affected by scaling.

3.2.2 Maximum Likelihood (ML) Device Classification

Herein MDA is considered for the RF-DNA classification and model development process, with Maximum Likelihood (ML) employed to determine decision boundaries for classification using equal priors and uniform costs [92]. This research considers identification as a classification problem, where the classifiers are built to determine a device's identity from its RF-DNA fingerprints using training/reference fingerprints and testing fingerprints. This is considered as a *one-to-many* comparison [19]. When examining the ML case, classification involves computing the Bayesian posterior probabilities from the classifier, for N_C a fingerprint F^ω is assigned to class ω_i if

$$P(\omega_i|F^\omega) > P(\omega_j|F^\omega), \forall j \neq i, \quad (3.16)$$

for $i \in \{1, 2, \dots, N_C\}$ devices [19]. The conditional probabilities for such problems are Bayesian in nature:

$$P(\omega_i|F^\omega) = \frac{P(F^\omega|\omega_i)P(\omega_i)}{P(F^\omega)}, \quad (3.17)$$

where the denominator is constant across ω_i for a given F^ω [19]; with equal priors for all classes, $P(\omega_i) = 1/N_D$. The likelihood is estimated through a Gaussian distribution:

$$P(F^\omega|\omega_i) = \frac{1}{(2\pi)^{n_{df}/2} |\Sigma|^{1/2}} \exp(\mathcal{F}_e), \quad (3.18)$$

with \mathcal{F}_e being a form of Mahalanobis distance:

$$\mathcal{F}_e = -\frac{1}{2}(F^\omega - \mu)^T \Sigma^{-1}(F^\omega - \mu), \quad (3.19)$$

for the sample mean, μ , and inverse covariance, Σ^{-1} , of the data with as implicit assumption of normality [19].

3.3 Learning Vector Quantization Family of Methods

Although the improved Generalized Relevance Learning Vector Quantization (GRLVQ) algorithm of Mendenhall, [244–247], is of primary interest herein due to its previous application to RF-DNA classification and verification in [51, 92, 100]. Beyond RF-DNA classification and verification, LVQ methods have seen a wide variety of applications, ranging from image analysis [244–246, 248], to disease detection [249]. To fully understand GRLVQI, one must necessarily understand the workings and philosophy of LVQ and the successive extensions to GRLVQ to further extend the LVQ family of algorithms.

Epistemologically, LVQ methods are neural networks. Broadly, there are three categories of neural network approaches: feedforward, recurrent, and self-organizing maps, with LVQ methods included in the last category [250]. This is conceptualized in the general taxonomy of Artificial Neural Networks (ANN) shown in Figure III-2, where ANN types and basic examples of their architectures, and how nodes and layers connects, are presented. Broadly, LVQ refers to a family of supervised neural learning approaches which learns input relevance with classification as part of its cost function [245, 250–254]. The LVQ family of methods includes various extensions and improvements from vector quantization (VQ) and the LVQ algorithms developed by Kohonen, [255–257].

Both VQ and LVQ are considered as neural network functions due to similarities in the iterative training approach used for VQ and LVQ prototype vectors, which are analogous to ANN hidden-layer nodes, the use of gradient descent for training and the non-linearity of the process [213]. Additionally LVQ can be seen as a *nearest neighbor* approach through the *nearest prototype vector* (PV) optimization process [258].

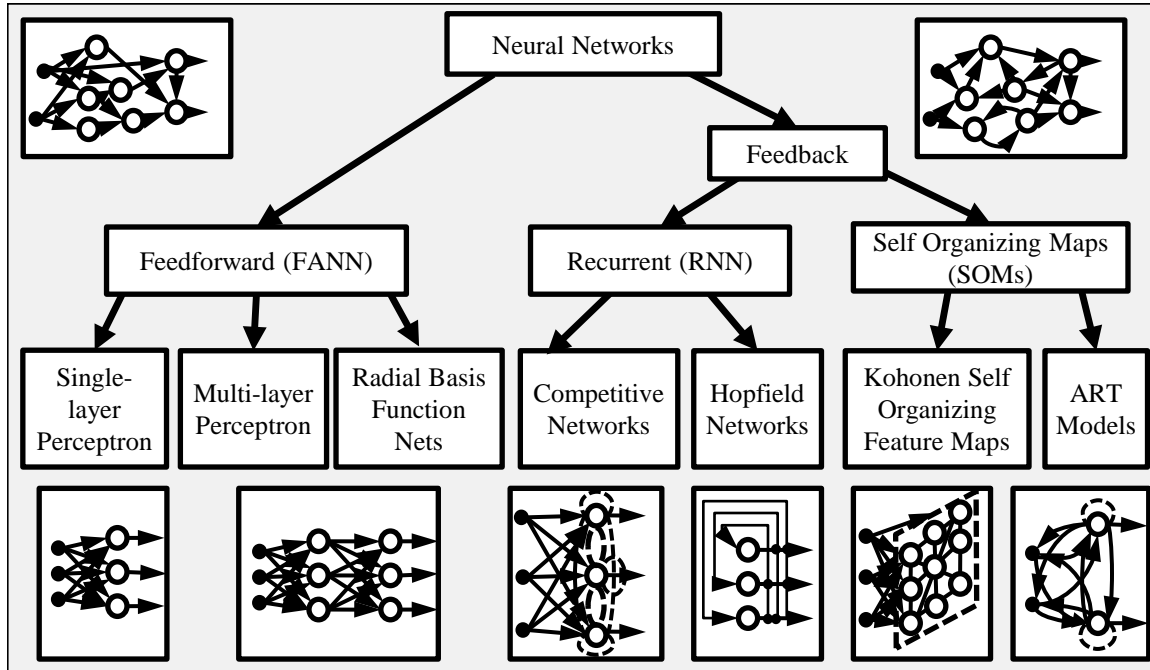


Figure III-2: General taxonomy of ANN approaches, adapted from [254] using the ANN families of [250, p. 368].

While PVs and hidden nodes appear analogous, a few distinctions exist between LVQ and ANN networks. Primarily, in LVQ, each PV is associated with a specific class resulting in LVQ methods being “winner take all” methods where one and only one PV will win for each exemplar [259–261]. Additionally, this also means that LVQ does not employ an output layer [262]. Therefore, LVQ could be considered as an ANN with no

explicit output layer and a winner take all hidden/output layer. These differences between ANNs and LVQ are conceptualized in Figure III-3.

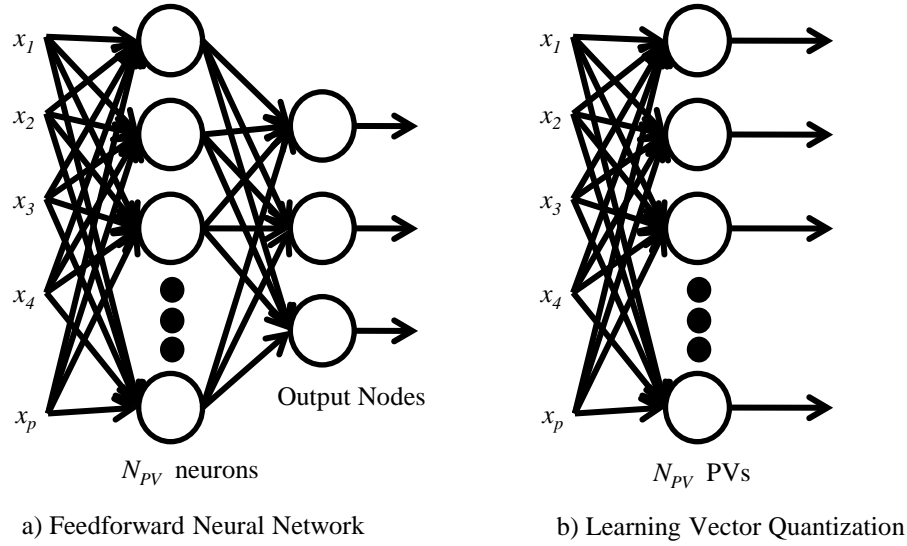


Figure III-3: Conceptualization of the differences between a) ANNs and b) LVQ networks, adapted from [250, 262].

For classification, a constraint exists where PVs must implicitly correspond to a true data class. Logically this implies that the number of PVs should be $N_{PV} \propto N_C$, hence if $N_C = 3$ then N_{PV} must be in multiples of 3. PVs are then initialized with random values and assigned to the corresponding classes, with PVs indexed $1, \dots, N_{PV}/N_C$ being associated with class 1 and so on. In operation, PVs are considered as organized en bloc, e.g. if $N_{PV} = 3$ for $N_C = 3$ classes, then $w_1(t)$ represent true class 1, $w_2(t)$ represent true class 2, and so on.

Classification of PVs to data exemplars is considered iteratively through a distance measure, nominally squared Euclidean distance. Conceptualized in Figure III-4 is the general process for LVQ variations, using the logic of LVQ2.1. In Figure III-4 we

are observing the closest in-class PV, w^J , and closest out-of-class PV, w^L , to the i^{th} data exemplar, x_i , based on the respective distances, d^L and d^J . Iteratively, PVs, w , are compared to a given training set exemplar and either a) moved closer to the corresponding same-class sample (for correctly classified PVs), and/or b) moved further away from the out-of-class sample (for incorrectly classified PVs). Depending on the LVQ variant and PVs strategy, a window can be incorporated to further restrict which PVs are updated.

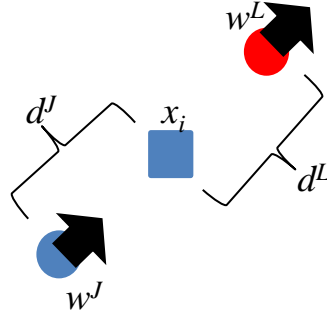


Figure III-4: LVQ prototype vector update conceptualization; adapted from [249].

3.3.1 Gradient Descents and LVQ

Gradient descents involve iteratively moving PVs, or nodes, appropriately towards or away from a given exemplar [216]. Followed appropriately, resultant PVs would accurately characterize the data with lower dimensionality [216]. The general definition of a linear gradient descent appears as

$$w(t + 1) = w(t) - \epsilon(t)\nabla C(w(t)), \quad (3.20)$$

where t is the training sample iteration number, $\epsilon(t)$ is a learning rate, $w(t)$ is a given PV, and $C(w(t))$ is a cost function and ∇ implying the gradient [216, 263]. Care must therefore be taken in specifying the learning rate, initializing the PVs and in selecting the cost function.

All LVQ methods follow a similar gradient descent based approach, as presented in (3.20), to move PVs towards or away from data as needed. LVQ methods typically differ only with respect to the cost function, update logic, and the inclusion of additional computational steps (e.g. relevance computations). Major variations are reflected through the addition of letters to the LVQ acronym, a brief taxonomy of major LVQ variations leading from LVQ to GRLVQI is presented in Table III-1. Kohonen first extended LVQ by creating variants (cf. LVQ2 and LVQ2.1) that improved the PV update strategy to updates involving both in-class and nearest out-of-class PVs [255]. Relevance LVQ (RLVQ) extends LVQ by incorporating a relevance weight for each data feature, which is learned during the training process [264]. GLVQ extends LVQ by improving class boundary approximations through the incorporation of a sigmoid cost function and the use of gradient (first derivative) descent [265]. Hammer and Villmann's [266] GRLVQ, combined the innovations of both GLVQ and RLVQ to create a GLVQ algorithm that learned the input dimension weights to provide relevance information regarding each feature. GRLVQ was then further extended through improvements resulting in the GRLVQI algorithm [244, 245]. A table describing the various versions of LVQ leading up to GRLVQI is provided in Table III-1. Other variations that divert from

LVQ in PV update approach, logic rules, algorithm formulations, and other methods are not considered herein. Such innovations include: LVQ4 [267], kernel LVQ variants [268, 269] and information theory based approaches [270]. Further extensions and philosophies of LVQ variations are documented in reviews, such as provided by Nova and Estevez [252], Kaski et al. [271], and Kaden et al. [258].

Table III-1: Major Variations in LVQ Family of Algorithms.

VERSION	VARIATION	REFERENCE
VQ	An unsupervised clustering ANN/gradient descent approach where PVs are moved towards data exemplars to create a feature space.	[255, 257, 272]
LVQ	A supervised clustering (classification) version of VQ which either pushes correctly classified PVs towards a given group and incorrectly classified PVs away. Includes Kohonen variants, in addition to LVQ2, LVQ2.1, and LVQ3	[256, 257]
GLVQ	A generalized form of LVQ, reference vectors are updated with a sigmoid used in the cost function/gradient descent	[265, 273]
RLVQ	LVQ modified with a gradient descent based input feature relevance computation	[264]
GRLVQ	A combination of the innovations in RLVQ and GLVQ. Incorporates 2 gradient descent operations. Weighting factors for inputs incorporated into the GLVQ method, permitting scaling of input dimension by relevance.	[266]
GRLVQI	GRLVQ with the following improvements: improved prototype update rule, improved prototype utilization, and a frequency based maximum input update strategy	[245–246]

3.3.1.1 Vector Quantization (VQ)

VQ and the Self Organizing Feature Map (SOFM) clustering method are approaches that aim to represent the input data, \mathbf{X} , as N_{PV} total PVs [216, 274]. VQ operates by iteratively selecting a random data exemplar and then using a gradient descent operation to move the nearest PVs towards the given exemplar [255, 272]. In operation, first N_{PV} must be selected and these PVs must then be initialized appropriately [255]. Similar to other clustering problems, it is non-trivial to decide on the number of PVs (N_{PV}) to be created [275–278]. However, some care must also be taken in initializing PVs for VQ. Logically, $N_{PV}/N_C > 1$ is of interest, and PVs initialized with identical values will yield dubious results; therefore PVs initialized as all zeros are a poor choice, and hence initializing with random values is seen in practice [255]. It is also helpful if the PVs and the data have the same dynamic range, therefore one reasonable solution would be to standardize the data, \mathbf{X} , and then use PVs from a random normal distribution [255].

After initializing the PVs, the distances between a given i^{th} exemplar and each of the $n = 1, \dots, N_{PV}$ PVs are computed to find the index of the PV associated with the minimum distance [255]. Nominally, squared Euclidean distances are used for the distance measure in VQ, with the cost function being the distance measure itself

$$d_n = C(w_n(t)) = (x_i - w_n)^2, \quad (3.21)$$

the PV associated with the minimum distance, $w_d(t)$, is then updated through the gradient descent process in (3.20). The chain rule, as described in Edwards and Penney [279] as

$$\frac{du(g)}{dv} = \frac{du}{dg} \frac{dg}{dv}, \quad (3.22)$$

where $u(g)$ is a function, u , of another function, g . Considering (3.22) with $u = (x_i - w_n)^2$ and $g = (x_i - w_n)$, one can compute the derivative for the squared Euclidean cost function. Following this formulation, the gradient of the cost function is computed as

$$\nabla C(w_d(t)) = -2(x_i - w_d(t)), \quad (3.23)$$

and is then used to update a given PV [255]. The scalar multiplier can be combined with the learning rate, and the VQ gradient descent operation is thus computed as,

$$w_d(t + 1) = w_d(t) + \epsilon(t)(x_i - w_d(t)), \quad (3.24)$$

which flips the sign of (3.20) due to the negation seen in the gradient.

3.3.1.2 Learning Vector Quantization (LVQ)

LVQ extends upon the concepts of VQ by creating essentially a supervised version of VQ to enable classification [253, 255, 257, 280]. Similar to VQ, N_{PV} PVs are defined and initialized appropriately with preference towards the PVs and the data sharing a similar dynamic range [255]. Thus instantiating random normal PVs and standardizing the input data is one common

In operation, LVQ begins similar to VQ where the distances between a given i^{th} exemplar and each PV is again computed per (3.20) [253, 255, 257, 280]. However, the gradient descent operation now depends on whether a correct classification was made or not. Here, when $w_d(t)$ is associated with the corresponding class of x_i , a correct classification was made. The gradient descent process of (3.20) for the i^{th} exemplar follows a Hebbian learning process [281],

$$w_d(t+1) = \begin{cases} w_d(t) + \epsilon(t)(x_i - w_d(t)) & \text{if } C_d = C_i \\ w_d(t) - \epsilon(t)(x_i - w_d(t)) & \text{if } C_d \neq C_i \end{cases}, \quad (3.25)$$

where conditions for correct and incorrectly classified PVs are both considered, with C_i being the class identity of the i^{th} exemplar and C_d being the class identity of the PV under consideration [255]. In (3.25), $C_d = C_i$ indicates a correctly classified exemplar and $C_d \neq C_i$ indicates an incorrectly classified exemplar [253, 255, 257, 280].

3.3.1.3 Learning Vector Quantization Improvements (LVQ2 and LVQ2.1)

Three general philosophies exist on improving LVQ, including 1) altering the update logic of (3.25), 2) incorporating additional gradient descents, and 3) changing the cost function. Kohonen [282] first proposed LVQ2 as an extension of LVQ logic that only updates PVs when they were appropriately close to a given exemplar. In LVQ2 [282], a window and various criteria are introduced. LVQ2 and LVQ2.1 are conceptualized via Figure III-5. LVQ2 extends the PV update logic in (3.25) where the two closest PVs to a given exemplar x_i are considered. PVs are updated if and only if (iff) 1) x_i falls within the window, 2) x_i belongs to K_L , and hence 3) the two nearest PVs are an in-class PV and out-of-class PV. In this process x_i lies within the window if

$$\min\left(\frac{d^L}{d^J}, \frac{d^J}{d^L}\right) > 1 - \iota, \quad (3.26)$$

where ι is a scale factor having a recommended value of approximately 0.35 [282].

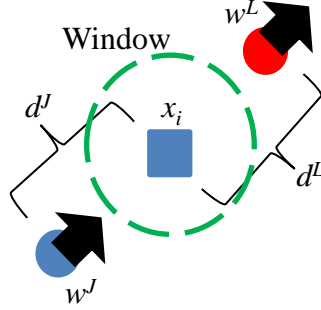


Figure III-5: Conceptualization of the LVQ2 and LVQ2.1 prototype vector update approach using the LVQ2.1 process; adapted from [282].

Kohonen [282] admitted that LVQ2 had various issues, e.g. computationally intensive and slow convergence, and therefore proposed a further variation in LVQ2.1. LVQ2.1 considers the basic LVQ algorithm with the LVQ2 logic, however the difference is that LVQ2.1 does not wait for the class of x_i to serendipitously match w^L and rather finds both of the nearest in-class PVs and nearest out-of-class PV to x_i [282].

LVQ2.1's PV update logic extends (3.25) where the in-class PV is moved toward the data exemplar,

$$w_n^J(t+1) = w_n^J(t) + \epsilon(t) (x_i - w_n^J(t)), \quad (3.27)$$

and the out-of-class PV is moved away from the data exemplar

$$w_n^L(t+1) = w_n^L(t) - \epsilon(t) (x_i - w_n^L(t)), \quad (3.28)$$

if x_i falls within the update window [282]. In many subsequent LVQ implementations, e.g. GLVQ and GRLVQ, the general logic of LVQ2.1 is followed for updating prototype

vectors. Additionally, one of the primary improvements seen in GRLVQI is an extension of the LVQ2.1 logic.

3.3.1.4 Relevance Learning Vector Quantization (RLVQ)

RLVQ was introduced by Bojer et al. [264] as an extension of LVQ that determines feature relevance during the classification process. Bojer et al. [264] recommend initializing the feature relevance weights ψ as a vector of length N_F with all values initially equal to $1/N_F$.

Per Hammer and Villmann [266] the RLVQ relevance update expression introduced by Bojer et al. [264] can be computed for each q^{th} data feature as a gradient descent,

$$\psi(t+1) = \psi(t) - \xi(t) \nabla C(\psi), \quad (3.29)$$

where ψ are scalar relevance values associated with a given data feature, and $\xi(t)$ is the relevance learning rate [264]. The distance from (3.21) for updating relevance rankings is considered, per [266], as

$$d_n = C(\psi) = \psi \cdot (x_i - w_n)^2. \quad (3.30)$$

The resultant relevance updates are thus updated for the q^{th} data feature via

$$\psi_q = \begin{cases} \psi_q - \xi(t) (x_{iq} - w_{nq}(t))^2 & \text{if } C_d = C_i \\ \psi_q + \xi(t) (x_{iq} - w_{nq}(t))^2 & \text{if } C_d \neq C_i \end{cases}, \quad (3.31)$$

with in-class and out-of-class considerations consistent with LVQ and (3.25). Per Hammer and Villmann [266], the RLVQ expression in (3.31) was formulated per the gradient descent. This formulation indicates that when the cost function changes, one

must necessarily change the ψ as well. The gradient descent operation and derivation for PV updates obviously do not change due to the inclusion of the scalar weighting term. Otherwise, the LVQ operation and logic of (3.25) do not change.

3.3.1.5 Generalized Learning Vector Quantization (GLVQ)

GLVQ extends LVQ through considering a sigmoidal cost function for the gradient descent in (3.20) rather than the linear cost function that produced the generic VQ gradient descent formulation of (3.24) [265]. The cost function considered in GLVQ algorithms is,

$$C = \sum_{m=1}^{N_{\text{Samples}}} f(\mu(x^m)), \quad (3.32)$$

at iteration t for N_{Samples} samples [245, 265]. The function $f(\mu(x^m))$ in (3.32) is a sigmoid function defined as

$$f(\mu(x^m)) = \frac{1}{1 + e^{-\mu(x^m)}}, \quad (3.33)$$

of the relative distance difference measure $\mu(x^m)$ [262].

In GLVQ, GRLVQ and GRLVQI, the relative distance difference measure is typically defined as

$$\mu(x^m) = \frac{(d^J - d^L)}{(d^J + d^L)}, \quad (3.34)$$

that appears related to the Soresen and Canberra distance metric, cf. [283, 284], with d^J and d^L being the respective squared Euclidean distances between the input sample x^m and the best matching in-class prototype vectors w^J , and best matching out-of-class

prototype vector w^L [245, 252, 265, 266]. The classification performance is inherently incorporated into (3.34) and, in operation, (3.34) is a normalized value between -1 and 1, which equates to a correct classification when $\mu(x^m) < 0$, a perfect classification (distance from in-class PV to exemplar approaches 0 while the distance from the out-of-class PV to the exemplar is large) when $\mu(x^m) = -1$, and incorrect classifications when $\mu(x^m) \geq 0$ [245, 265]. Due to the direction of correct and incorrect classification in (3.34), minimization is desirable to improve classification performance. This computation is also termed a “difference-over-sum” normalization or “normalized difference” and sees application in other domains, cf. [285–289]. The general concept also bears similarity to an alternative LVQ PV update representation of $w_n(t+1) = (1 - s(t)\epsilon(t))w_n(t) + s(t)\epsilon(t)x_i$, where $s(t)$ has a dynamic range spanning +1 for correct classifications and -1 for incorrect classifications [280]. Appendix B further examines the characteristics of (3.34).

One requirement of the distance measures used for d^J and d^L is that they must be differentiable for the gradient descent operation [290]. This makes logical sense, as a gradient is the first derivative. The nominal distance measure used in GLVQ is the same squared Euclidean distance seen in (3.21), however the derivation is complicated due to the formulation of (3.32)–(3.34). After computing the derivative associated with the gradient descent, PVs are computed via

$$\begin{aligned}
w^J(t+1) &= w^J(t) + \frac{4\epsilon(t)(\partial f / \partial \mu(x^m))d^L}{(d^J + d^L)^2} (x^m - w^J), \\
w^K(t+1) &= w^K(t) - \frac{4\epsilon(t)(\partial f / \partial \mu(x^m))d^J}{(d^J + d^L)^2} (x^m - w^K),
\end{aligned} \tag{3.35}$$

which are, respectively, the in-class and out-of-class updates for the winning PVs [245].

3.3.1.6 Generalized Relevance Learning Vector Quantization (GRLVQ)

GRLVQ involves the combination of the relevance method of RLVQ applied to GLVQ [266]. Therefore, the GLVQ cost function in (3.32) is extended in GRLVQ as,

$$C = \sum_{m=1}^{N_{\text{samples}}} \psi_q f(\mu(x^m)), \tag{3.36}$$

where ψ is again the relevance [245, 266]. The relevance approach of (3.31) changes to

$$\psi_q = \psi_q - \xi(t) f' |_{\mu(x^m)} \left(\frac{d^K}{(d_\lambda^J + d_\lambda^K)^2} (x^m - w^J)^2 - \frac{d^J}{(d_\lambda^J + d_\lambda^K)^2} (x^m - w^K)^2 \right), \tag{3.37}$$

because GRLVQ employs the cost function and PV updates of GLVQ [266]. Hammer and Villmann also recommend scaling relevance factors to ensure $\|\psi\|_1 = 1$ to avoid instabilities [266]. Consistent with the process of GLVQ, for GRLVQ PVs are computed via

$$\begin{aligned}
w^J(t+1) &= w^J(t) + \frac{4\epsilon(t)(\partial f / \partial \mu(x^m))d^L}{(d^J + d^L)^2} \Psi \cdot (x^m - w^J), \\
w^K(t+1) &= w^K(t) - \frac{4\epsilon(t)(\partial f / \partial \mu(x^m))d^J}{(d^J + d^L)^2} \Psi \cdot (x^m - w^K),
\end{aligned} \tag{3.38}$$

which is the formulation in (3.35) with the inclusion of the relevance term [266]. Additionally, some variants of GRLVQ incorporate different learning rates for in-class and out-of-class updates, as seen in [291].

3.3.1.7 Improved Generalized Relevance Learning Vector Quantization (GRLVQI)

Mendenhall [244] noted various issues in GLRVQ, including divergence due to unconditional updating of winning out-of-class prototype vectors. Mendenhall [244], and Mendenhall and Merenyi [245, 246] developed the GRLVQI algorithm to rectify these issues by improving the GRLVQ process in three ways: an improved update strategy, an improved learning rule to avoid classifier divergence, and improved prototype utilization.

(a) Improved Update Strategy

GRLVQI first has a new update strategy that adds a scalar time decay term, τ , to the miscalculation measure in (3.34) becoming

$$\mu(x^m) = \tau \frac{(d^J - d^K)}{(d^J + d^K)}, \tag{3.39}$$

which also implied, per [244–246, 292], that

$$f'(\mu(x^m), \tau) = f(\mu(x^m), \tau)(1 - f(\mu(x^m), \tau)) . \quad (3.40)$$

Since, τ is defined as a scalar, per Section 2.3.2.1 of [244], it therefore does not affect the derivation process related to the gradient descent operations in GLVQ and GRLV and the underlying framework of these algorithms is left intact.

(b) **Improved Learning Rule**

The improved GRLVQ algorithm incorporates a new learning rule by specifying that only the out-of-class prototype vector should be updated if a misclassification occurs [244]. Therefore, the improved GRLVQ algorithm update rule is as presented in Table III-2.

Table III-2: Improved GRLVQ Update Rule of Mendenhall [244]

Condition	Rule
Misclassification	<ul style="list-style-type: none"> • Move in-class prototype vector towards exemplar • Move out-of-class prototype vector away from exemplar
Correct Classification	<ul style="list-style-type: none"> • Move in-class prototype vector towards exemplar

(c) **Improved Prototype Utilization**

Mendenhall [244], and Mendenhall and Merenyi [245, 246] applied the ‘conscience’ learning of DeSieno [293] to in-class PV selection. The underlying philosophy is to discourage (bias) frequent PV winners from winning too often and encourage selection of infrequently selected PVs [245]. This is accomplished by computing the “frequency” of winning for the winning PV

$$F_{new}^P = F_{old}^P + \beta(1.0 - F_{old}^P) , \quad (3.41)$$

and adjusting the frequency in the non-winning PVs via,

$$F_{new}^P = F_{old}^P + \beta(0.0 - F_{old}^P) , \quad (3.42)$$

where β is a user defined parameter to control the updating [245]. The winning PV selection approach is also updated from (3.30) by subtracting β ,

$$d_{Bias} = d^P - \beta^P , \quad (3.43)$$

where d^P is either the in-class or out-of-class distance and β^P is defined as

$$B^P = \gamma \left(\frac{1}{P} - F_{old}^P \right) , \quad (3.44)$$

where γ is a scaling on the amount of bias, P indicates the PV number and F_{old}^P is the frequency [244].

3.3.1.8 Operational Settings for LVQ and GRLVOI

Determining operational settings for LVQ algorithms is a balance between science and art [244]. Although PV initialization is known to affect the classifier development in all LVQ variants [267], little has been published about LVQ algorithmic settings beyond specific guidelines for specific applications. A few considerations must be made, an appropriate learning rate needs to be specified for the gradient descent, PVs should be initialized to unique and appropriate vectors, and the appropriate number of PVs should be initialized.

(a) Learning Rates

Determining an appropriate learning rate $\epsilon(t)$ involves some consideration of the LVQ algorithm, architecture, and the data. Some general learning rate guidance exists

for differing algorithms. Selecting a learning rate for the gradient descent approach involves some work; too high of a learning rate introduces oscillations and possibly divergence, too low of a learning rate results in a slow convergence [216, pp. 312-313]. As mentioned in Mendenhall [244], there are “no hard-and-fast rules” in selecting learning rates and its selection is part of the “art of classifier design.” Per Strickert et al. [291], a general hierarchy relating learning rate $\epsilon(t)$ and relevance rate $\xi(t)$ includes $0 \leq \xi(t) \leq \epsilon(t) \leq 1$, assuming unscaled learning rates. In general the guidance of Kohonen [255] should be followed, where $\epsilon(t)$ is specified as a monotonically decreasing sequence of scalar values $0 \leq \epsilon(t) \leq 1$. Ideally, the monotonically decreasing term will either reach zero as an optimal solution is found or be stationary. This is logical because a decreasing/stationary learning rate avoids large movement within the data space as a solution becomes more refined.

Various general recommendations exist for LVQ learning rates, for instance Kohonen et al. [280] recommend learning rates of $\epsilon(t) \leq 0.1$ for LVQ. Although Bojer et al. [264] suggest initializing both the LVQ and relevance learning rates at $\epsilon(t) = 0.1$, they also employed different settings with RLVQ, such as $\epsilon(t) = 0.005$ and $\xi(t) = 0.05$ for a large mushroom dataset. Lim et al. [294] additionally suggested a default of $\epsilon(t) = 0.03$ for LVQ.

GLVQ and GRLVQ are more complicated algorithms and deserve further considerations. For general sigmoidal networks, which could feasibly include GLVQ, Duda, Hart and Stork [216, pp. 312-313] posit that a learning rate of $\epsilon(t) = 0.1$ is often adequate for initialization. This mirrors the general recommendations for LVQ learning

rate initializations. For one dataset, Hammer and Villmann [266] suggested a learning rate of $\epsilon(t) = 0.1$ and relevance rate of $\xi(t) = 0.01$; they further discussed the importance of the relevance rate being initialized smaller than the learning rate since the relevance is updated each iteration.

GRLVQI is a further more complicated algorithm, with three learning rates to select: PV learning rate, relevance learning rate, and conscience learning rates. Care must be taken since the interaction of these three learning rates is obviously complex and learning rates too high in magnitude could logically introduce instability and wild movements. In GRLVQI, there are two gradient descent learning rates, the PV learning rate $\epsilon(t)$ and the relevance rate $\xi(t)$, and two conscience parameters (γ and β) to consider, as seen Table III-3. Prior work determined operational settings for GRLVQI empirically, with Mendenhall [244], Mendenhall Table 3.3 [244], and Bischoff [295] recommending the $\epsilon(t)$ and $\xi(t)$ values presented in Table III-3. Bischoff et al. [295] empirically determined their recommended values by sampling each exemplar six times in random order during each of the N_{TS} total Training Step iterations. Additionally, the learning parameters in Table III-3 are learning schedules, which provide learning rate values depending on the quantity of training steps GRLVQI is employing. Table III-3 are implemented due to performance benefits seen and discussed in Mendenhall [244]. Table III-3 are not decaying values and thus learning rates are stationary during the specified training steps.

Table III-3: Nominal GRLVQ and GRLVQI Learning Parameter Learn Schedule.

NUMBER OF TRAINING STEPS N_{TS} (THOUSANDS)	GRLVQ PARAMETERS		CONSCIENCE PARAMETERS		REFERENCE
	$\epsilon(t)$	$\xi(t)$	γ	β	
$0 < N_{TS} \leq 400$	0.005	0.025	2	0.35	[244]
$400 < N_{TS} \leq 800$	0.0025	0.0125	2	0.3	
$800 < N_{TS} \leq 1200$	0.001	0.005	2	0.225	
$1200 < N_{TS} \leq 1600$	0.0005	0.0025	2	0.125	
$0 < N_{TS} \leq 500$	0.005	0.005	2.5	0.35	[295]
$0.5 < TS \leq 1$	0.0025	0.0025	2.0	0.30	
$1 < TS \leq 1.5$	0.001	0.001	1.5	0.225	
$1.5 < TS \leq 2$	0.0005	0.0005	1.0	0.125	
$2 < TS \leq 2.5$	0.00025	0.00025	0.75	0.1	

(b) Number of Prototype Vectors

Additionally, little is written on the appropriate number of PVs to instantiate. Kangas et al. [296] indicated that no unique solution existed for this task, but provided guidance (albeit without examples or proofs) that proportions to the number of samples in classes could be a wrong strategy. Georgiou [262] posited that more resolution is offered by increasing the number of PVs. Mendenhall [244] notes that generalization bound methods such as Gaussian complexity [297] can be used to determine the upper bound on the number of PVs to instantiate. One could expect that too many PVs would lead to over fitting, as mentioned in [298], and that too few would lead to poor classification performance. Therefore, selecting the appropriate number of PVs is of interest, despite little being written on it.

A general restriction in LVQ algorithms exists that the data must contain at least two classes and that there must be at least one PV per class [299]. However, further guidance on the number of neurons/prototype vectors to initialize is rarely mentioned in publications. Kohonen merely mentions that the optimal number of PVs is generally not proportional to the prior probability of classes [282]. Additionally, it was suggested that PVs could be deleted during the learning process [282]. But, no general framework was presented to suggest the appropriate number of PVs to initialize.

(c) **Prototype Vector Initialization**

A final consideration in LVQ network initialization is the proper initialization of the PV vectors themselves. Basic PV initialization approaches include using data sampling distribution [244], extreme points in the data [300], borders between classes [296], or random values [266, 301]. Additionally, some literature suggests initializing PVs using k-means to find cluster centers [267, 282], self-organizing maps [282] or by finding the means of each class [282]. However, employing *k*-means or self-organizing maps is akin to a fusion process of an unsupervised classifier feeding into a supervised classifier and *k*-means is iterative and not computationally free. PV initialization was a concern of Mendenhall [244], resulting in the addition of conscience parameters in the GRLVQI algorithm.

Logically, the key aspect of any PV initialization process is that the PVs and data exist in the same space; obviously, PVs should be initialized to be near the data dynamic range or else valuable iterations will be spent moving towards the data. Two obvious and

logical choices exist for proper PV initialization: 1) initializing PVs to normal random values and standardizing the data to have a dynamic range comparable to the random values, and 2) initializing PVs to random values in the data space. Herein, and consistent with [51], PVs are initialized with random normal values with the data standardized via standard score normalization [302],

$$z = \frac{x - \mu}{\sigma}, \quad (3.45)$$

where μ is the mean of a given data vector and σ its standard deviation.

(d) Number of Training Iterations

Similar to the issues of PV initialization, learning rate initialization, and selecting the number of PVs, very little appears in literature on selecting N_{TS} . For LVQ, Kohonen [255] recommends $N_{TS} = 500 \times N_{PV}$ as a general rule. Literature recommends various numbers of iterations, including $150 \leq N_{TS} \leq 600$ [303], $500 \leq N_{TS} \leq 2,500$ [295], $N_{TS} = 1200$ [51], a maximum of $N_{TS} = 10,000$ [255], and $400K \leq N_{TS} \leq 1.6M$ [244].

Reising [51] adopted an approach where multiple iterations were employed and then the best models were selected. Such an approach is consistent with the method employed by Gage [304] for ANN training. Gage [304] adopted Welch's method [305] for convergence to determine when to stop training and how many training epochs to use. Rather than find steady-state operating conditions, one looks for a stable operating point where volatility has decreased [304, 305]. Hence, this is a visual approach to determine where data "converges" [304, 305]. Similar to the approach of Gage [304], Reising [51] computed the GRLVQI model at each iteration and then determined which model offered

the best performance. The best performing model was then used for subsequent analysis and for comparison against the sequestered test set.

3.4 Device Classification and Verification Methodology

For model development, classification accuracy is the standard performance metric used for the RF-DNA problems; however, it is analyzed in different ways depending on the task at hand (classification/model building or verification). Historically, c.f. [18, 89, 92, 113, 224], the Air Force Institute of Technology's (AFIT's) RF-DNA development has considered *Device Classification* as a one-to-many “looks most like?” assessment, and *Device ID Verification* as a one-to-one “looks how much like?” assessment.” In operation, this involves classification being used for model development using the library at hand with verification examined when new devices attempt to claim the identity of a known device. These concepts extend from the biometrics concepts of *enrollment*, collecting templates from users; *verification*, validating a user's identity through comparison with that user's template; and *identification*, recognizing a user by searching the entire database [6].

3.4.1 Classification Performance

RF-DNA classification performance generally considers evaluation of training, testing, and validation (in GRLVQI) performance of classifier models. Both the MDA/ML and the GRLVQI processes were applied using a full-dimensional ($N_F = 729$) RF-DNA feature set extracted from ZigBee emissions collected to support results in [91]. Classification results are presented in Figure III-6 displaying that MDA/ML overall

outperforms GRLVQI while both show a general pattern of high classification accuracy for high SNR with relatively lower classification accuracy at lower SNR. Classification results for Z-wave devices are similarly presented in Figure III-7. Comparing performance to these baseline results is one general approach used to evaluate algorithmic performance throughout this research.

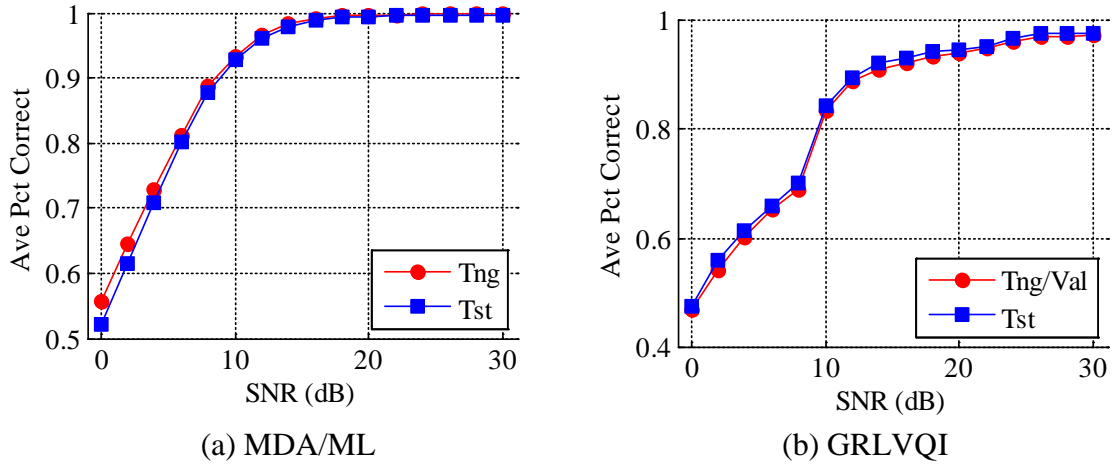


Figure III-6: ZigBee Full-dimensional Baseline Classification Results

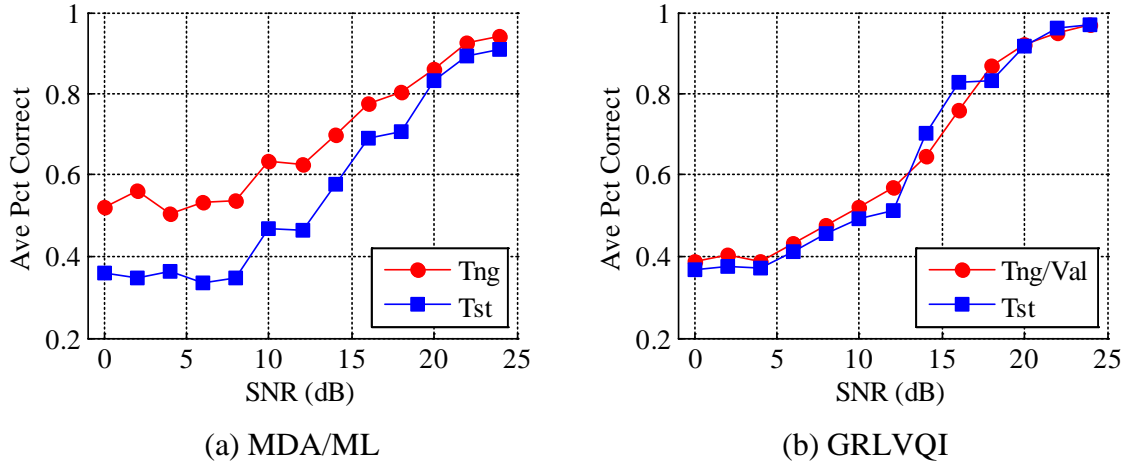


Figure III-7: Z-Wave Full-dimensional Baseline Classification Results

3.4.1.1 Classification Performance Assessment: Gain Trade-Offs

One basic approach employed to compare classification performance between competing algorithms, or performance of a given algorithm for various settings, is relative performance gain G_{SNR} . Consistent with prior RF-DNA works [51], performance gain G_{SNR} is defined herein as the reduction in required SNR, expressed in dB, for the two methods under consideration to achieve a given average percentage of correct classification (%C). This definition is depicted in Figure III-8 for MDA/ML and GRLVQI testing performance of %C = 90%. When comparing MDA/ML and GRLVQI, we examine performance at a nominal, arbitrary operating point of %C = 90%. As indicated in Figure III-6 MDA/ML requires $SNR = 8.68$ dB (TNG) and $SNR = 8.99$ dB (TST), while GRLVQI requires $SNR = 12.92$ dB (TNG) and $SNR = 12.39$ dB (TST) to achieve the same performance. Thus, for ZigBee MDA/ML is superior and provides a gain of $G_{SNR} = 3.4$ dB (TST) relative to GRLVQI.

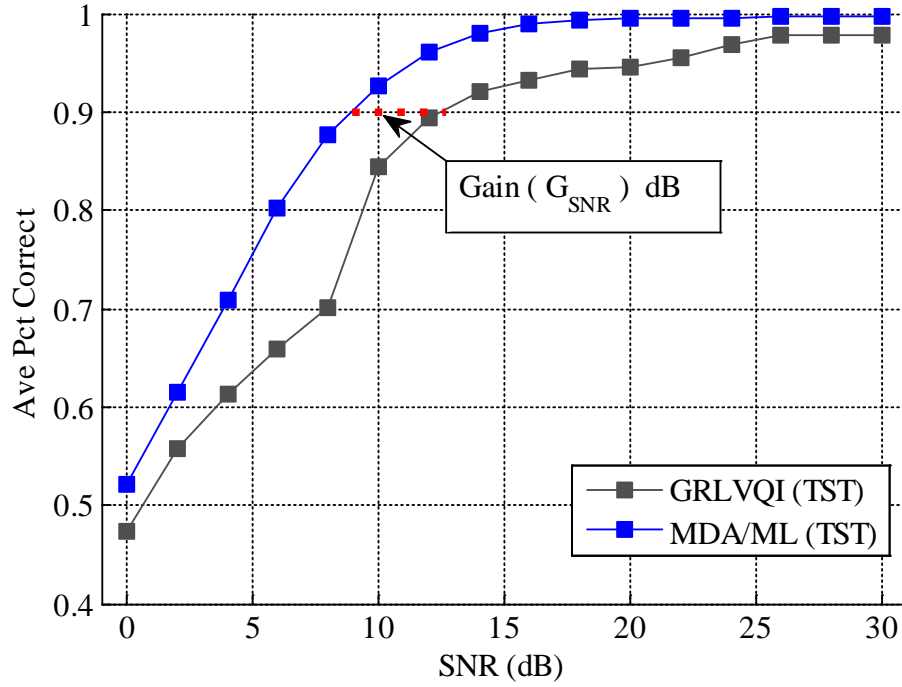


Figure III-8: Gain Trade Off Example for MDA/ML (TST) and GRLVQI (TST) for ZigBee.

If one similarly considered TST results in Figure III-7 for Z-Wave, GRLVQI is seen to be superior and yields a relative MDA/ML gain of $G_{SNR} = +3.32$ dB (TST). Therefore, when considering classification performance, GRLVQI is a superior classifier for Z-Wave, while MDA/ML was a superior classifier for ZigBee.

3.4.1.2 Classification Performance Assessment: Relative Accuracy Percentage (RAP)

To facilitate broader comparison of %C versus SNR performance, a Relative Accuracy Percentage (RAP) metric was introduced in Bihl et al. [135]. The RAP is generated by first computing the Area Under Classification Curve (AUCC) values for each method being compared. This is done using a trapezoidal approximation, with a

given method's estimated $AUCC_{M(i)}$ being in the denominator and the baseline $AUCC_{Base}$ being in the numerator

$$RAP = AUCC_{M(i)} / AUCC_{Base} . \quad (3.46)$$

According to (3.46), RAP provides the fraction of $AUCC_{M(i)}$ with respect to $AUCC_{Base}$ and enables both 1) a comparison for methods that do not achieve the arbitrary $\%C \geq 90\%$ performance benchmark, and 2) a comparison reflecting performance across all SNR levels. Interpreting RAP values is also intuitive, with 1) $RAP < 1.0$ indicating that the method under consideration achieves lower $\%C$ than the baseline method, 2) $RAP = 1.0$ indicating that the method under consideration achieves $\%C$ performance comparable to the baseline, and 3) $RAP > 1.0$ indicating that the method under consideration exceeds baseline $\%C$ performance.

For the ZigBee results in Figure III-6 with MDA/ML serving as the baseline, $AUCC_{Base} = 27.18$ (TST), $AUCC_{GRLVQI} = 25.24$ (TST) and $RAP = 0.93$ indicating that MDA/ML performs better across all operating points than GRLVQI. For Z-Wave results in Figure III-7, MDA/ML $AUCC_{Base} = 13.32$ (TST) and $AUCC_{GRLVQI} = 15.06$ (TST), yielding $RAP = 1.13$ which indicates that GRLVQI performs better across all operating points when compared to MDA/ML for Z-Wave.

3.4.2 Device ID Verification

In essence, device ID verification is a form of conditional classification which considers a *one-to-one* comparison of a device's actual identity with its claimed identity [19]. This approach approximates a trained and tested classifier when examining

possibly new, previously unseen data. A device is considered to be authentic when the posterior probability

$$P(\omega|F^{New}) \geq t , \quad (3.47)$$

with F^{New} being a newly observed RF-DNA fingerprint; ω , the class the device claims the identify of; and t being a decision threshold [19]. Device ID verification performance is then evaluated by plotting Receiver Operating Characteristic (ROC) curves over various decision thresholds [19].

3.4.2.1 Verification Performance Assessment: ROC Curves

Consistent with [89] two error conditions are evaluated: False Verify Reject (FVR), for *rogue* devices, and False Reject Rate (FRR), for *authorized* devices. FVR and FRR are respectively evaluated against either True Verify Rate (TVR) or True Rejection Rate (TRR) to generate ROC performance curves [89], consistent with the general ROC methodology of [306]. The equal error rate (EER) point on these ROC-like curves is either 1-TVRR for *authorized* or 1-TRR for *rogue*. Consistent with prior research, e.g. [89], verification performance will be evaluated as %*Authorized* or %*Rogue Rejected* at 90% TVR/TRR and 10% FVR/FRR.

3.4.2.2 Baseline Verification Performance

When examining verification performance at 18dB, Figure III-9 and Figure III-10 for authentic vs rogue devices, MDA/ML appears to achieve perfect verification, Figure III-9a and Figure III-10a, while GRLVQI presents considerably lower verification performance. Therefore improving GRLVQI to make it a viable RF-DNA algorithm is of

major importance to both ensure multiple competing classifier methods are vetted for future RF-DNA research and to understand what could be leading to this deficiency.

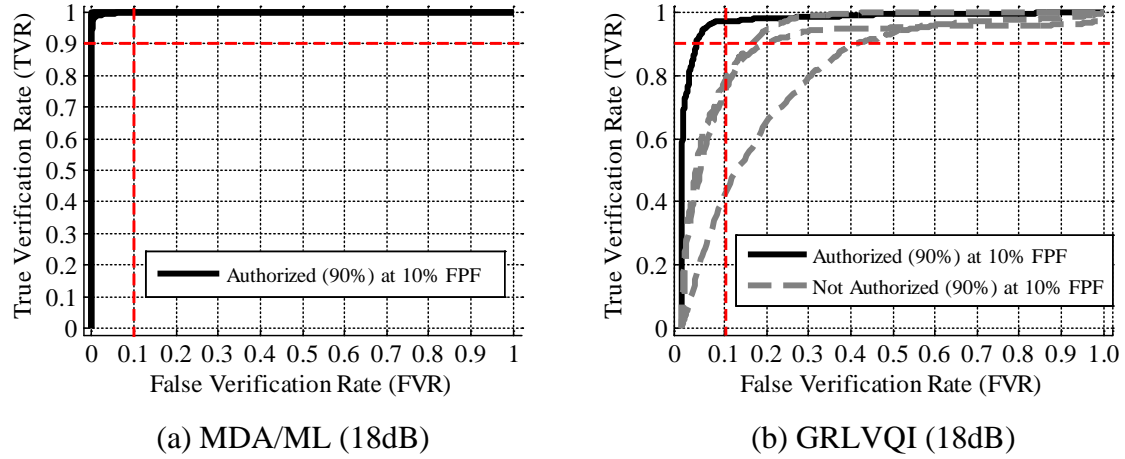


Figure III-9: ZigBee MDA/ML and GRLVQI full-dimensionality authorized device verification results baseline

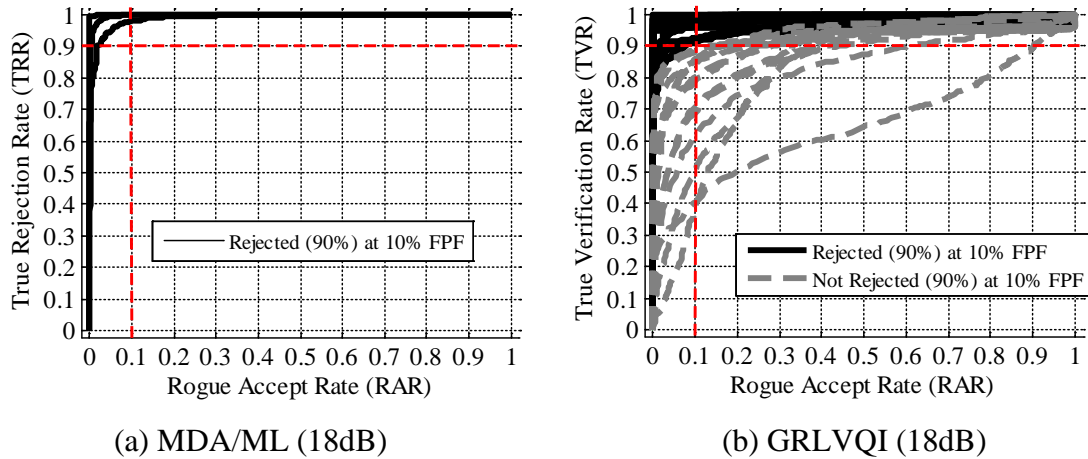


Figure III-10: ZigBee MDA/ML and GRLVQI full-dimensionality rogue device verification results baseline

Figure III-11 presents verification results for Z-wave devices at 20dB using the MDA/ML classifier, Figure III-11a, and the GRLVQI classifier, Figure III-11b. Since

only $N_{Dev}=3$ devices are in the Z-Wave dataset, only authorized device results are presented. Although Z-Wave fingerprints were associated with higher GRLVQI classification performance, here one can see that verification performance is better with MDA/ML.

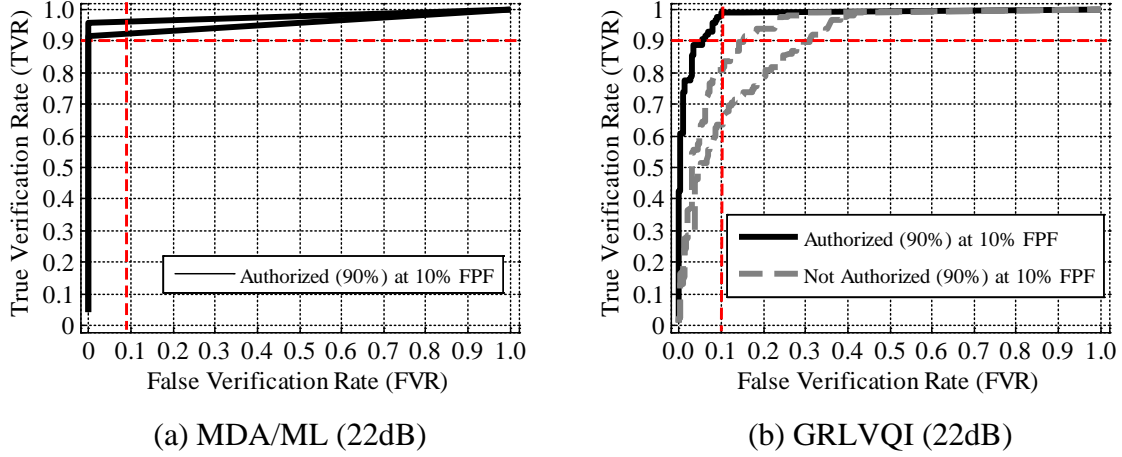


Figure III-11: Z-Wave MDA/ML and GRLVQI full-dimensionality authorized device verification results baseline

3.4.3 MDA/ML and GRLVQI Baseline Results

Overall classification results for MDA/ML and GRLVQI using both ZigBee and Z-Wave RF-DNA feature sets are presented in Table III-4. The relative RAP and Gain metrics in Table III-4, with MDA/ML serving as the baseline method (highlighted in grey), illustrate that MDA/ML generally outperforms GRLVQI for both ZigBee RF-DNA classification, while Z-Wave achieves generally better classification performance using the GRLVQI classifier.

Table III-4: Baseline Classification Results.

DEVICE	ALGORITHM	DATA SET	AUCC	SNR AT %C = 90%	RELATIVE MDA/ML (TST) RAP	RELATIVE MDA/ML (TST) GAIN (G _{SNR})
ZigBee	MDA/ML	Training	27.39	8.68 dB	1.01	+0.31
		Testing	27.18	8.99 dB	1.00	0.00
	GRLVQI	Training	24.99	12.92 dB	0.92	-3.93
		Testing	25.24	12.39 dB	0.93	-3.4
Z-Wave	MDA/ML	Training	16.39	21.23 dB	1.23	+1.68
		Testing	13.32	22.91 dB	1.00	0.00
	GRLVQI	Training	15.23	19.19 dB	1.14	+3.72
		Testing	15.06	19.59 dB	1.13	+3.32

For the ZigBee results in Figure III-6 with MDA/ML serving as the baseline, $AUCC_{Base} = 27.18$ (TST), $AUCC_{GRLVQI} = 25.24$ (TST) and $RAP = 0.93$ indicating that MDA/ML performs better across all operating points than GRLVQI. For Z-Wave results in Figure III-7, MDA/ML $AUCC_{Base} = 13.32$ (TST) and $AUCC_{GRLVQI} = 15.06$ (TST), yielding $RAP = 1.13$ which indicates that GRLVQI performs better across all operating points when compared to MDA/ML for Z-Wave.

Authorized and Rogue device verification results, for ZigBee, are presented in Table III-5, for selected SNR operating points. Table III-5 illustrates that MDA/ML generally achieves higher verification accuracy at lower SNR than GRLVQI. Consistent with the ZigBee results, authorized verification results for Z-Wave, are presented in Table III-6 for selected SNR operating points, which again illustrates that MDA/ML generally achieves higher verification accuracy at lower SNR than GRLVQI.

Table III-5: ZigBee Baseline Device ID Verification Results.

Algorithm	Operating SNR (dB)	Verification Scenario	Verification Accuracy (%)
MDA/ML	10	TVR (%)	100
	10	RRR (%)	100
	14	TVR (%)	100
	14	RRR (%)	100
	18	TVR (%)	100
	18	RRR (%)	100
	20	TVR (%)	100
	20	RRR (%)	100
	22	TVR (%)	100
	22	RRR (%)	100
GRLVQI	10	TVR (%)	0
	10	RRR (%)	8.33
	14	TVR (%)	25
	14	RRR (%)	47.22
	18	TVR (%)	25
	18	RRR (%)	63.88
	22	TVR (%)	50
	22	RRR (%)	75

Table III-6: Z-Wave Baseline Device ID Verification Results.

Algorithm	Operating SNR (dB)	Verification Scenario	Verification Accuracy (%)
MDA/ML	10	TVR (%)	0
	14	TVR (%)	66
	18	TVR (%)	100
	22	TVR (%)	100
GRLVQI	10	TVR (%)	0
	14	TVR (%)	0
	18	TVR (%)	0
	22	TVR (%)	66

3.4.4 MDA versus GRLVQI in RF-DNA Research

While MDA/ML consistently out-performs GRLVQI in both classification and verification tasks for ZigBee, Z-Wave devices saw better classification performance using GRLVQI and better verification performance using MDA/ML. It is therefore advantageous to consider further research in GRLVQI developments with emphasis towards RF-DNA applications because MDA/ML has known deficiencies in certain contexts.

Firstly, based on the criteria in (3.10), MDA is limited when the number of classes exceeds the number of available features, a possible situation if many devices were considered in a real world setting where ZigBee networks can contain up to 65,000 devices [39]. However, it should be noted that 1) all pattern recognition methods have performance issues (accuracy and computationally) as the number of classes grows into the 10s to 100s (let alone 1000s) as seen in the literature on “highly multiclass” problems,

c.f. [307–314], and 2) linear methods, such as MDA, are commonly employed in “highly multiclass” problems due to their computational advantages, c.f. [310, 315, 316].

Secondly, as seen in Reising [51] and the Z-Wave dataset results in Figure III-7 and Figure III-11, GRLVQI does outperform MDA/ML in some RF Fingerprinting applications. Thirdly, data distributions and particularly bimodality can cause issues in MDA with respect to finding the best discriminant direction, as seen in [317], which are logically possible given the many varied applications of RF-DNA. Therefore, ample motivation exists for improving and furthering the understanding of GRLVQI and applying such improvements for further RF-DNA research.

IV. Dimensional Reduction Analysis

Men dig up and search through much earth to find gold.

–HERACLITUS, 535BC – 475BC

Given large volumes of data being collected in many domains, e.g. big data [318–327], the primary challenge becomes selecting relevant data features for a given task. Dimensional Reduction Analysis (DRA) is therefore of interest to select salient subsets of a dataset for analysis.

4.1 Introduction

As Ruskin states in [328], “For all books are divisible into two classes: the books of the hour, and the books of all time,” thus, indicating that relevance and importance is critical. Hayek similarly notes in [329] that many problems can be reduced to logic “...if we possess all the relevant information, if we can start out from a given system of preferences and if we command complete knowledge of available means.” Many datasets contain more data than necessary for reliable classification which, inherently, becomes a problem that can be addressed using DRA to improve performance after discarding non-salient features [330]. One concept in feature selection is that feature salience is linked to dependence on class labels [331], therefore feature selection methods that result from classifier model development (termed *post-classification*) and methods that consider the distribution of data with respect to a class label vector (e.g. Analysis of Variance (ANOVA) based F-test) are of particular interest.

This chapter is organized as follows. Section 4.2 presents, develops and discusses various DRA methods, with Section 4.2.1 discussion *pre-classification* DRA methods, Section 4.2.2 discussion *post-classification* DRA methods, Section 4.2.3 developing MDA based DRA methods, Section 4.2.4 discussing DRA fusion, Section 4.2.5 discussing Random DRA as a baseline method, and Section 4.2.6 discussion dimensionality assessment methods. Section 4.3 then considers Multiple Discriminant Analysis (MDA) models and ZigBee RF-DNA features to assess various DRA methods for device discrimination, including both *Device Classification* (1 vs. N_C assessment) and *Device ID Verification* (1 vs. 1 assessment).

4.2 Dimensional Reduction Analysis Methods

DRA can consist of many processes and actions; at the highest level, DRA is considered to embody three aspects: *dimensionality assessment* (*qualitative* versus *quantitative*), *feature selection* versus *feature extraction*, and *pre-classification* versus *post-classification*. The following describe higher level aspects of DRA:

1. *Pre-classification* versus *post-classification*: The distinction between *pre-classification* and *post-classification* DRA involves where in the overall pattern recognition process the DRA is performed. *Pre-classification* DRA involves any method performed *a priori* of any classification step, e.g. input data distribution-based methods, while *post-classification* DRA is performed *a posteriori* of the classification step and includes information from the classifier on feature relevance, e.g. MDA loadings [237, pp. 394-429] [242],

Artificial Neural Network – Signal to Noise Ratio (ANN-SNR) feature screening [330] and Relevance Learning Vector Quantization (RLVQ) methods [51]. *Pre-classification* DRA is also known as *filter* methods and *post-classification* is also known by the term *embedded* or *wrapper* methods [332, 333]. Since *pre-classification* DRA is not directly tied to classifier performance it does not necessarily improve classifier performance, as seen in [334].

2. *Feature selection* versus *feature extraction*: consistent with [213, 335], *feature selection* involves selecting subsets of existing features through *pre-classification* or *post-classification* feature relevance rankings, while *feature extraction* involves a data transformation into either a lower dimensional space or a transformed space, e.g. the RF Distinct Native Attribute (RF-DNA) Fingerprinting Process itself, MDA, or Principal Component Analysis (PCA). Feature selection is relevant throughout many domains including multivariate statistics to manufacturing [336].
3. *Dimensionality Assessment*: DRA also involves an operator judgment on the amount of data to retain. Both *qualitatively* and *quantitatively* dimensionality assessment methods can be used. *Quantitative* dimensionality assessment computationally determines the amount of data or what features to retain, whereas *qualitative* dimensionality assessment involves subjective selection of the quantity of features. In some application, subject matter expertise can be leveraged for *qualitative* dimensionality assessment [89, 91, 113] where

subjective amounts of features were retained. Quantitative dimensionality assessment methods are considered here using heuristics on data covariance matrix eigenvalues, MDA loadings, and test statistic p-values.

Excluding the RF-DNA Fingerprinting feature extraction process itself, described in Section II, prior DRA research for RF-DNA has considered three feature selection methods: A) a *pre-classification* distribution-based two-sample Kolmogorov-Smirnov goodness-of-fit test (KS-test), B) a *post-classification* GRLVQI feature relevance rankings process [91], and C) the *post-classification* Random Forest feature relevance rankings process [134]. While all three approaches have seen success in RF-DNA applications, logically DRA methods associated with classification, e.g. *post-classification*, should be associated with improved classification performance.

Of particular interest to this research were methods that could be used to

1. improve and expand the RF-DNA DRA foundation by improving the understanding of the KS-test DRA algorithm, which involves understanding the appropriate use of p-values and test statistics for feature relevance ranking,
2. extend the distribution-based one-way ANOVA F-statistic method to RF-DNA,
3. compare and contrast dimensionality assessment approaches,
4. aid development of an MDA-based DRA algorithm,
5. compare with GRLVQI feature relevance ranking, and

6. aid development of DRA fusion approaches to combine multiple feature relevance ranking approaches.

4.2.1 Distribution Based Feature Selection DRA

Distribution-based pre-classification feature selection for RF-DNA considers either data feature distributions with respect to class membership or data feature distributions against other features. Both approaches are considered herein using the two-sample KS-test and the F-statistic. Additionally, of particular interest is understanding whether test statistic values or probabilities (p -values) from the tests are best for achieving reliable feature relevance ranking.

4.2.1.1 Two Sample Kolmogorov-Smirnov (KS) Test

The KS-test was codified by Massey [337] based on independent contributions by Kolmogorov [338] and Smirnov [339]. The KS-test is a distribution-based goodness-of-fit process for comparing the distribution of a given sample vector (\mathbf{x}_1) with a given reference distribution [337]. The two sample KS-test is an extension that quantifies differences in cumulative distribution functions for two sample vectors (\mathbf{x}_1 and \mathbf{x}_2) using a test statistic of the form,

$$KS = \max(|F_1(\mathbf{x}) - F_2(\mathbf{x})|) \quad (4.1)$$

where $F_1(\mathbf{x})$ is the proportion of \mathbf{x}_1 values less than or equal to \mathbf{x} , $F_2(\mathbf{x})$ is the proportion of \mathbf{x}_2 values less than or equal to \mathbf{x} , and KS is the computed test statistic value [337, 340, 341]. With the test statistic, KS , being the maximum difference between the curves, if \mathbf{x}_1 and \mathbf{x}_2 come from the same distribution, the value of KS converges to zero. Higher

values of KS indicate different distributions while lower KS values indicate similar distributions [337; 340, pp. 344-385].

For determining p -values, the underlying KS-test null hypothesis is that \mathbf{x}_1 and \mathbf{x}_2 are from the same distribution and the alternative hypothesis that they are from different distributions [337, 340]. For the KS-test, data degrees of freedom (DoF) and the null distribution are used to compute p -values, with p -values ranging from 0 to 1 [340]. Additionally, KS-test p -values can identically equal 0 [340]. Although not mentioned in [91, 134, 241] and largely automated in practice, the process for computing approximated KS-test p -values is rather involved and requires first computing

$$\varsigma = \max \left(\left[\sqrt{N_e} + 0.12 + \frac{0.11}{\sqrt{N_e}} \right] KS, 0 \right), \quad (4.2)$$

where KS is the KS-test statistic value from (4.1) and

$$N_e = \frac{N_1 N_2}{N_1 + N_2}, \quad (4.3)$$

which represents the Harmonic mean [283] of the number of observations in Group 1 (N_1) and Group 2 (N_2) [342, pp. 623-628]. To compute the KS-test p -value, the following function is used

$$p_{est} = 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 \varsigma^2}, \quad (4.4)$$

with the final approximation of the p -value then computed as

$$p = \min(\max(p_{est}, 0), 1), \quad (4.5)$$

where the \min and \max functions ensure the estimate is bounded between 0 and 1 [337; 342, pp. 623-628; 343–345] .

Feature selection using the two sample KS-test was first proposed by Nechval [346] in 1988 for image processing and, prior to Dubendorfer [91], the KS-test saw limited DRA application with only one additional citation [347]. For RF-DNA DRA, KS-test p -values have seen many applications [89, 113, 134, 348]. For DRA, the KS-test is implemented pairwise in each feature by classes, where one should logically seek \mathbf{x}_1 and \mathbf{x}_2 from different distributions to avoid redundancy [113, 121]. For multiple classes, pairwise KS-test p -values are computed for each feature and then summed [91].

The formulation of the KS-test DRA algorithm in Figure IV-1 is based on Patel's [134] work and was revised here to include both A) the logical inequality of $i \neq j$ to ensure it is clear that only non-identical vectors are compared, and B) the correct inclusion of the test statistic from which the p -value is computed. The algorithm iteratively considers each feature via a pairwise comparison of the feature per class.

<p>Algorithm 1 KS-Test for Feature Selection</p> <pre> for Each feature $v = 1 \rightarrow N_F$ do for $i = 1 \rightarrow N_C$ classes do for $j = 1 \rightarrow N_C$ classes do if $i \neq j$ do \mathbf{x}_i = observations from class i, variable v \mathbf{x}_j = observations from class j, variable v $KS = \max(F_i(\mathbf{x}_i) - F_j(\mathbf{x}_j))$ $p(v) = p(v) + p(KS, DoF)$ end if end for end for end for </pre>
--

Figure IV-1: p -value KS-test Feature Selection Algorithm as adapted from Patel [134] and modified herein.

Figure IV-2 presents resultant summed p -values for ZigBee features using the algorithm in Figure IV-1. Results in this figure are consistent with observations made in [113, 121], i.e. phase (ϕ) features (indices 244 to 486) are collectively the most relevant (smaller p -values) when compared to amplitude (a) features (indices 1 to 243) and frequency (f) features (indices 487 to 729). However, it is evident in Figure IV-2 that a majority of features have very low (less than 0.1) summed p -values which may result in low feature selection resolution due to minute differences between relevance ranking values.

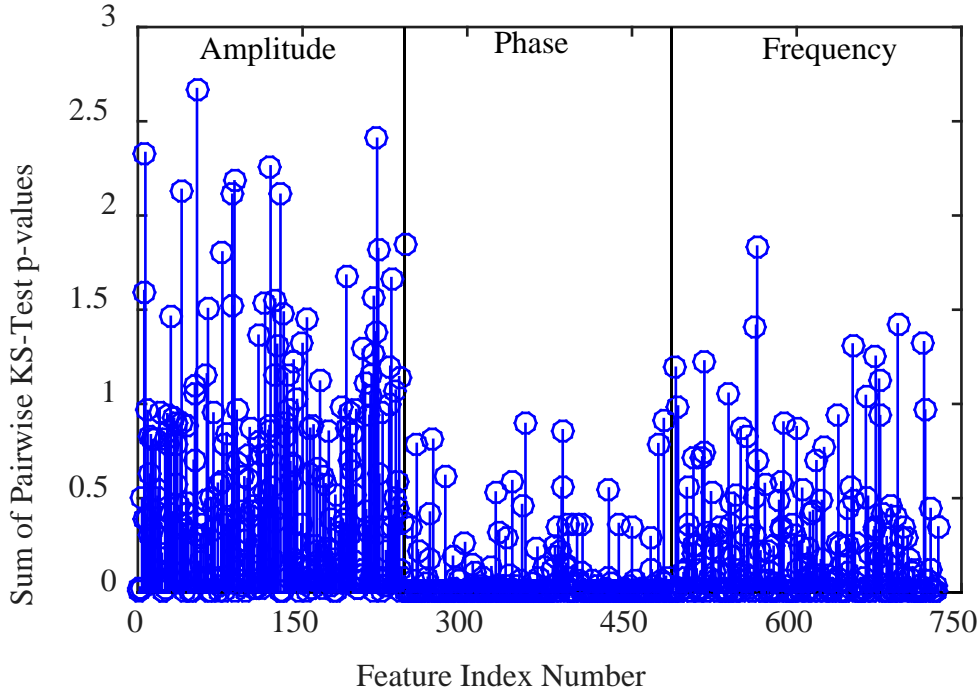


Figure IV-2: Sum of p -values from pairwise KS-test on ZigBee training observations using a full-dimensional ($N_F = 729$) feature set at SNR = 10 dB [89, 113]. Lower values indicate greater feature difference and potentially greater relevance.

Figure IV-3 presents the corresponding mean test statistic values for the p -values seen in Figure IV-2. Again, as in Figure IV-2, Figure IV-3 shows that phase (ϕ) features are most relevant (higher test statistic values). Incidentally, the p -values in Figure IV-2 trend toward zero while the test statistic values in Figure IV-3 do not trend to any single value.

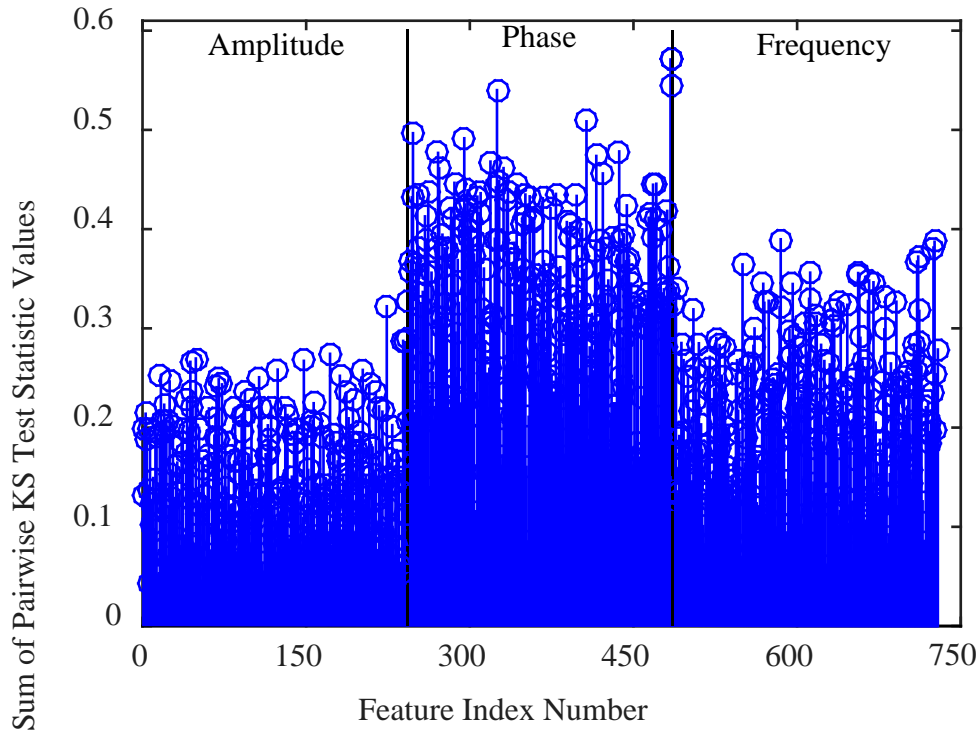


Figure IV-3: Mean of test statistic values from pairwise KS-test on ZigBee training observations using a full-dimensional ($N_F = 729$) feature set at SNR = 10 dB. Higher values indicate more different (and possibly more relevant) features.

4.2.1.2 One Way Analysis of Variance F-Statistic

Although previously unexplored for RF-DNA feature selection, feature ranking by F-statistic values from one-way ANOVA was first reported by Habbema and Hermans

[349] and has seen further application in medical [350], education data analysis [351], and other DRA applications [333, 352]. The underlying premise of F-statistic based DRA involves selecting features that provide a good relationship to the class membership, with the process echoing Hall and Smith's [353] advice that "a good predictor set should contain features highly correlated with the target class distinction, and yet uncorrelated with each other."

ANOVA considers a linear model which expresses the relationships between parameters as

$$Y = XB + \varepsilon , \quad (4.6)$$

where Y is a continuous response variable (each feature herein), X is an input variable (categorical vector of class identities herein), B are the solved parameters, and ε is a vector of iid assumed errors [302, 354, 355]. ANOVA employs the linear model in (4.6) to understand variability in observations through sum of squares computations of the observation from their mean and sum of squares associated from observational groups [302].

The F-test is a heuristic used to compute the significance of an ANOVA relationship, and is defined as

$$F_0 = \frac{MS_{Model}}{MSE} , \quad (4.7)$$

where MS_{model} is the mean square for a given general linear model between X and Y , and MSE is the mean squared error in a computed linear ANOVA model [302]. Traditionally, for ANOVA problems, p -values are computed from the F-test and used to

determine if a relationship is significant or not for the null hypothesis that there is no relationship between X and Y [302]. When considered as a feature selection problem, higher values of F_0 are taken to indicate that a feature is more likely to be useful in discriminating between classes [350]. To compute the p -value, the F -distribution is used, which has a probability density function,

$$f(x|u, v) = \frac{\Gamma\left(\frac{u+v}{2}\right) \left(\frac{u}{v}\right)^{\frac{u}{2}} x^{\left(\frac{u}{2}\right)-1}}{\Gamma\left(\frac{u}{2}\right) \Gamma\left(\frac{v}{2}\right) \left[\left(\frac{u}{v}\right)x + 1\right]^{(u+v)/2}} \quad (4.8)$$

with u and v being the respective Degrees of Freedom (DOF) for the numerator and denominator terms in (4.7) [302]. For RF-DNA application, u is the DOF due to groups ($N_c - 1$) and v is the DOF due to the number of observations ($N_{TNG} - u - 1$). Figure IV-4 presents the F -distribution computed for the entirety of the ZigBee training data, with $u = 3$ and $v = 4796$. The x-axis is in units of F-statistic value, as computed by (4.7), and the y-axis is the f-distribution value, as computed by (4.8) [302]. P -values are then computed by finding the area under the curve (AUC) at a given F-statistic value; these p -values are either one-sided (upper or lower tail) or two sided (both the upper and lower tail) [302]. For illustrative purposes, a two-sided test is used as this is what was used in practice. Further discussion of one-sided or two-sided test can be found in Montgomery and Runger [302].

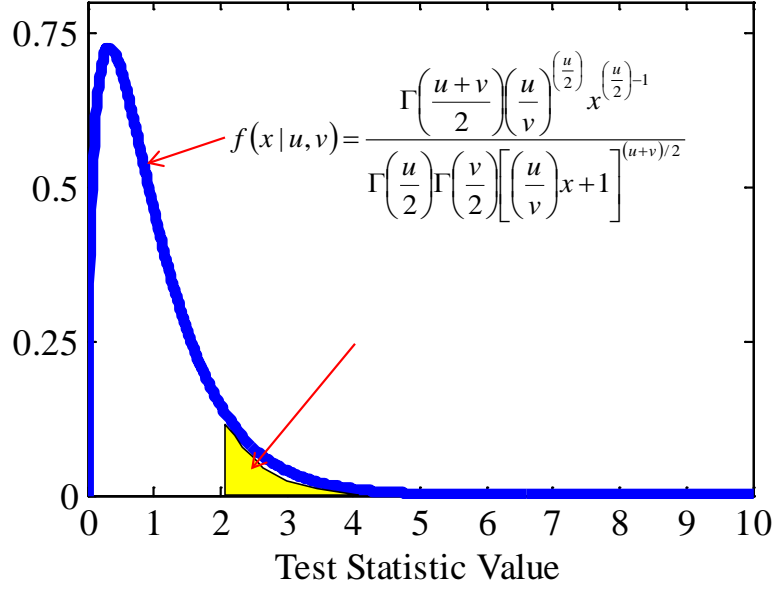


Figure IV-4: Example p -value computation from test statistics using an F-distribution.

Figure IV-5 presents an algorithm for feature relevance ranking using a one-way ANOVA F-test. Here, both test statistics and p -values are computed for each feature of the training data with respect to a corresponding class vector since [349–351] employed test statistics, and not the p -values, for feature relevance ranking.

Algorithm 2 F-Test Feature Selection Algorithm
for Each feature $i = 1 \rightarrow N_F$ do \mathbf{x}_i = observations from class i , variable i \mathbf{y} = vector of class identification $F\text{-test stat} = MS_{Model}/MS_{Error}$ $p(i) = p(F\text{-test stat}, DoF)$ end for

Figure IV-5: One way ANOVA F-test Feature Relevance Ranking Algorithm.

Figure IV-6 and Figure IV-7 present the test statistic and p -values, respectively, at $SNR = 10$ dB after employing Algorithm 2 on the ZigBee RF-DNA data. Consistent with

the KS-test, smaller p -values in Figure IV-7 are again considered as more relevant. Comparing Figure IV-6 and Figure IV-7, here one can see that both test statistics and p -values indicate that phase features are more relevant; however, one can also see that the p -values trend towards zero while test statistic values do not.

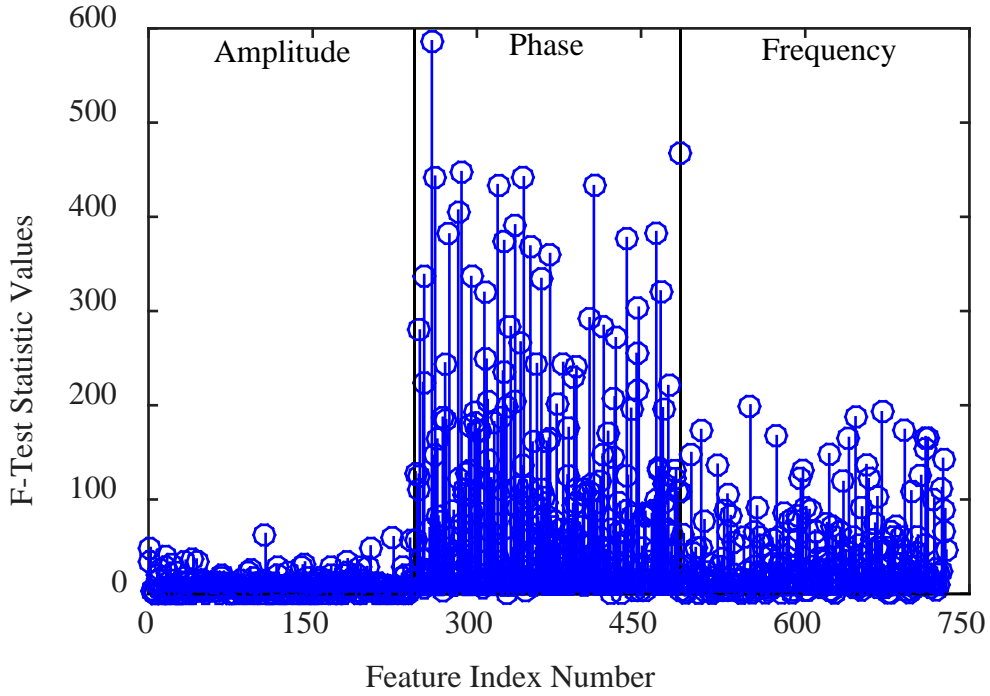


Figure IV-6: Test statistic values from F-test on ZigBee training observations using a full-dimensional ($N_F = 729$) feature set at SNR = 10 dB. Lower values indicate greater feature difference and potentially greater relevance.

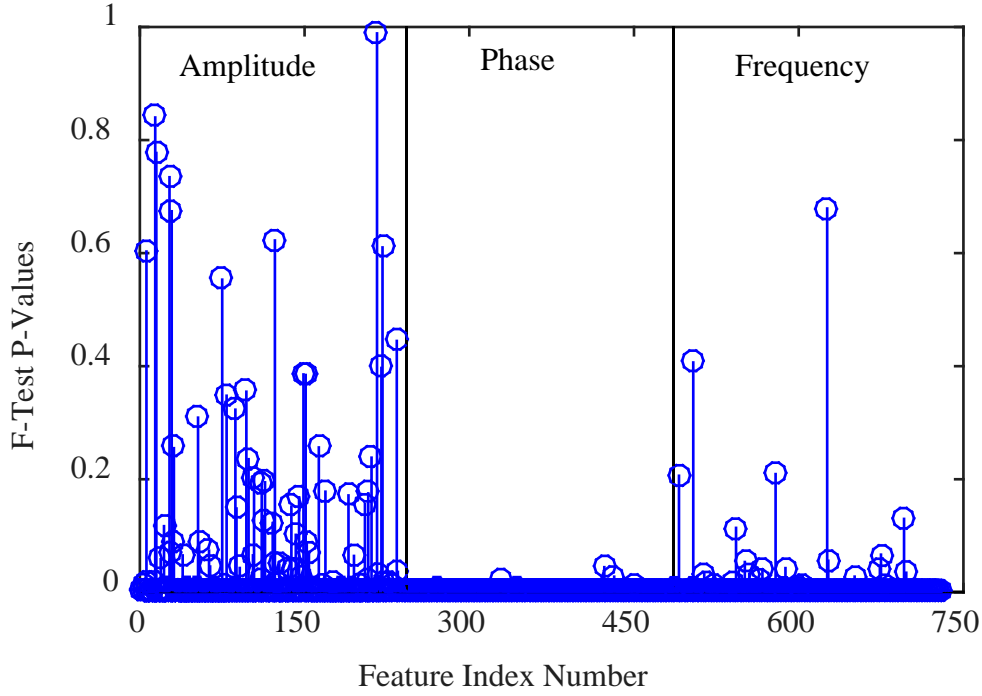


Figure IV-7: p -values from F-test on ZigBee training observations using a full-dimensional ($N_F = 729$) feature set at SNR = 10 dB. Lower values indicate greater feature difference and potentially greater relevance.

4.2.1.3 Test Statistic versus P -values for Feature Relevance Ranking

Test statistic values are commonly converted to p -values (probabilities) to assess significance [302]. P -values are generally considered as the smallest level at which an observed test statistic value is significant [356]. However, the appropriate use and the general appropriateness of p -values in statistics are associated with much debate. This is inherently related to the meaning of a p -value [357]. For feature relevance ranking, various studies consider p -values, c.f. [89, 113, 121, 358–361], and many backward and forward selection methods employ p -values for feature selection [362, 363]. KS-test p -values were also used by Wendt et al. [364] to compare similarities of distributions

between different foundries. However, others advocate the use of the test statistic itself [349, 351, 365].

Due to this disagreement in literature, an understanding of the use of p -values and test statistic values is needed. To facilitate this, a philosophical understanding of p -value and test statistics is first formulated, then a short description of the relative steps required to compute KS-test and F-test p -values, this is followed by an empirical understanding of p -values and test statistic values for DRA.

(a) **General Understanding of P-value Use and Misuse**

Essentially, a p -value is a reflection of a computed test statistic value given a probability distribution and for a specific null hypothesis [366]. When computed, the p -values indicate the probability of observing a given result given the reference distribution and the specified null hypothesis [367, 368]. Hence a p -value is only meaningful in the context of a given scenario [369], and to compute any p -value one necessarily needs the following quantities: a hypothesis test, data degrees of freedom, a reference probability distribution, a test statistic result, and a hierarchy of possible outcomes [367]. However, these are not always stated in feature relevance ranking applications, c.f. [89, 113, 121, 241, 370], and thus resultant p -value results are often presented out of context.

While test statistic values and p -values largely move in opposite directions (smaller p -values indicate larger test statistic values), the mapping is rarely linear and is associated with various properties of the reference distribution. Test statistics are often ratios of data dependent quantities while p -values refer to the probability of getting that

value which involves assumptions with respect to a distribution. Various issues therefore exist when using p -values for feature relevance ranking as noted by Cord et al. [365]. When interpreting p -values, differences in p -values can result from differences in effects sizes and/or differences in standard errors [371], and thus using p -values as a quantifiable value is considered a logical fallacy of the transposed conditional [372]. P -values are additionally viewed as imprecise and debate exists on whether approximate p -values are more useful than exact values [373].

Additionally, using p -values for feature relevance ranking appears akin to issues mentioned in Anderson et al. [374] where p -value magnitudes were shown to offer possibly erroneously interpretation of effect size. Other problems exist in that small p -values can be computed due to either low variability or large sample sizes [374]. For example, Kitbumrungrat [375] considered MDA as a classifier and presented feature relevance ranking values for an MDA-based DRA method, F-test, and p -values; while the p -values were all essentially equal, the other methods presented different relevant ranking values for each feature.

The larger question also exists on whether p -values are appropriate for feature relevance ranking; this particularly revolves around the issue of treating p -values as exacts when p -values of similar magnitude are essentially equivalent [369]. While one can point to many feature selection methods, such as forward/backward/stepwise regression, as using p -values for feature selection [354], using p -values for feature relevance ranking is not without controversy, c.f. [365, 376].

Some disagreement also exists in statistics literature on if it is appropriate to even use p -values for traditional hypothesis testing purposes, e.g. [357, 369, 377–390], with some journals even refusing to publish p -values from hypothesis tests, e.g. *Epidemiology* [391] and *Basic and Applied Social Psychology (BASP)* [377]. While some of this debate involves debates between Bayesian and Frequentists statisticians [392], further issues involve the incorrect application of p -values, as Senn [393] stated, “ p -values are a practical success, but a critical failure,” and issues relating to sample-to-sample p -value variability and the influence of sample size [369].

Summation and many other methods used to combine p -values may present some difficulties due to an implicit assumption that p -values are the result of independent tests. How to properly combine p -values is another issue and a variety of methods for differing conditions therefore exist, c.f. [394–403]. However, in prior RF-DNA applications, c.f. [89, 113, 121], summed p -values were not directly interpreted as probabilities, thus the chance for misinterpretation may not exist. Although, many of the steps listed in Sections 4.2.1.1 and 4.2.1.2 to compute either KS-test or F-test p -values are automated, these are implicit steps that cannot be ignored when employing a process. Additionally, by considering the steps needed to compute their respective p -values, we can conceptualize the issues that exist in p -value feature relevance ranking in the KS-test and F-test.

In summary, the various issues related to p -values for DRA include:

1. Resolution is lost in the mapping from the test statistic to the (typically) nonlinear p -value.

2. P -values are imprecise [373].
3. Computing p -values involves an implicit distributional assumption whereas test statistics are often only ratios.
4. That p -values frequently converge to zero for large quantities of samples [369].
5. An additional and unnecessary computation is required in looking up the associated p -value for a given test statistic, hypothesis test, degrees of freedom and distribution.
6. Fundamentally, p -values indicate statistical significance, but nothing about the magnitude of that statistical significance [404–406].
7. Prior to computing test statistic values, one is not making an explicit distributional assumption, but one must make a distributional assumption when computing a p -value. An example, the experimentally computed F -test statistic value in (4.7) is merely a ratio of sums of squares. While terming (4.7) an “ F -test statistic” does imply an F -distribution, until one formalizes a hypothesis test and computes the p -values, no distributional assumption has been made since there are no distributional assumptions with general linear models prior to these steps [407]. Therefore, test statistic values are generally ratios, but do not indicate any underlying inferences, or significance, of these values until they are tied to a hypothesis test and reference distribution.

(b) **P-value Versus Test Statistic Feature Relevance Ranking**

Beyond literature references regarding p -values, it is useful to empirically evaluate the p -values for feature relevance ranking. This is considered below for both the KS-test and F-test on the ZigBee RF-DNA Fingerprint data, and further in Appendix C on general academic datasets. As seen in Figure IV-4, the resulting p -value from a given test statistic involves firstly an additional computational step and secondly a nonlinear mapping. As one can visualize, the AUC will nonlinearly vary as a given test statistic may linearly vary, inherently making comparison, ranking, and interpretation more difficulty. Additionally, F-test p -values may not offer comparison of features from multiple datasets since the underlying probability distribution changes as the degrees of freedom change.

To examine the distributions of the p -values and test statistic values for the F-test and KS-test, histograms of unit area, using the same bin centers and bin widths, are used. Figure IV-8 presents summed p -values from the KS-test, while Figure IV-9 presented mean test statistic values from the KS-test. Four operating points, $SNR = [0, 10, 18, 30]$ dB are used in both Figure IV-8 and Figure IV-9. Overall, both Figure IV-8 and Figure IV-9 illustrate that features become more statistically significant in the KS-test as noise diminished with p -values approaching 0 as the underlying null hypothesis is rejected. However, conditions exist where all features could be viewed as significant if only p -values feature ranking were used. For instance, at $SNR = 10$ dB two features have a summed p -value equal to exactly 0, and at $SNR = 30$ dB, 99.7% of the features are in the first bin (centered at 0.0108) with 12% of the features having a p -value exactly equal to 0

and thus of equivalently relevant. This issue of resolution exists even at $\text{SNR} = 0$ dB, where a large number of features have very low p -values.

While feature relevance resolution was lost when using p -values, as seen in Figure IV-8, resolution is not lost when using test statistic values, Figure IV-9. The result in Figure IV-9 thus illustrates that KS-test statistic values offer a more refined and consistent approach for finding and selecting features which is not overwhelmed by the numerous p -value issues as described in Section 4.2.1.3 and visualized in Figure IV-8.

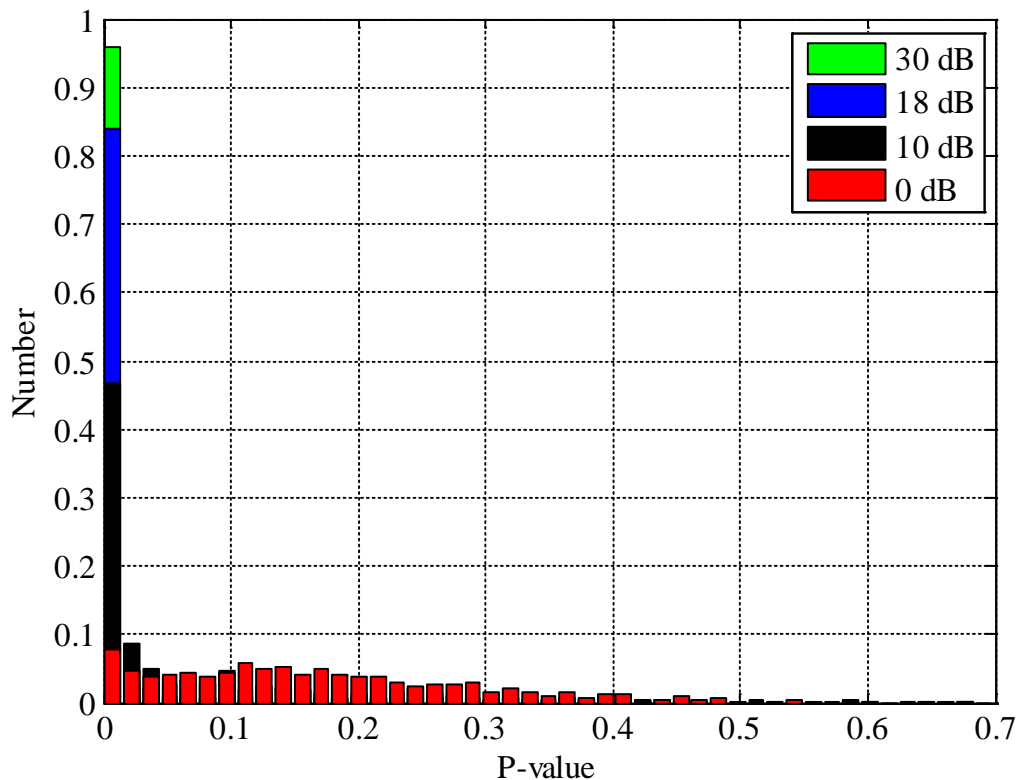


Figure IV-8: Normalized histogram of *summed* pairwise KS-test p -values using a full-dimensional ($N_F = 729$) ZigBee TNG feature set for varying $\text{SNR} = [0, 10, 18, 30]$ dB.

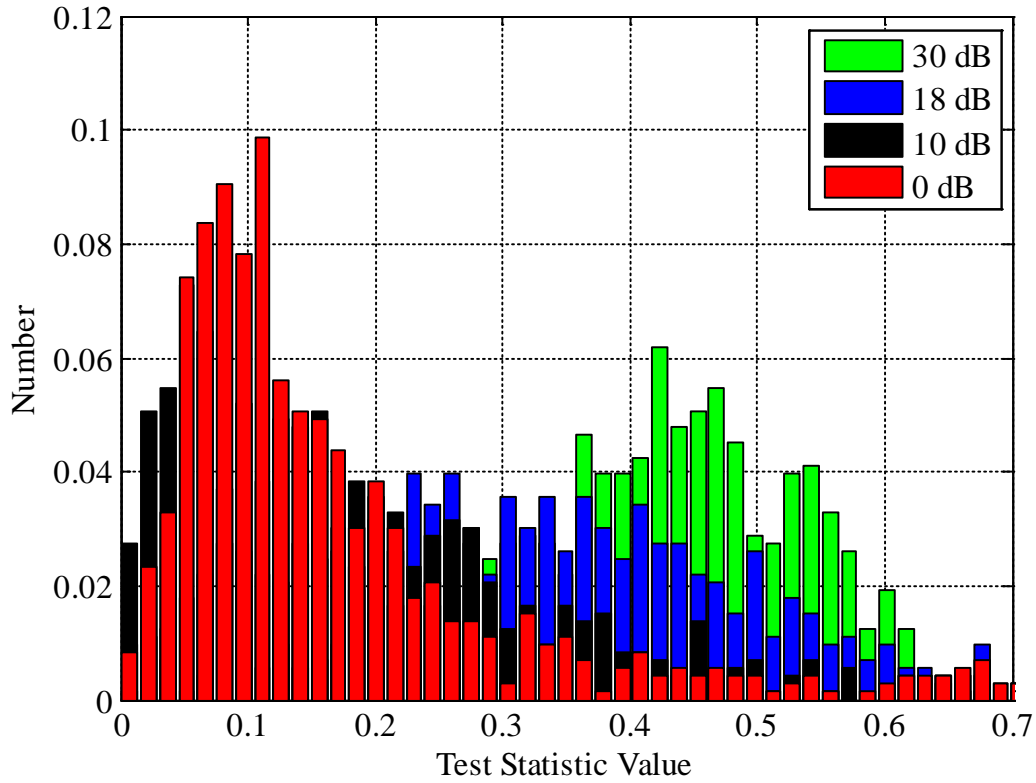


Figure IV-9: Normalized histograms of *mean* pairwise KS-test statistic values using a full-dimensional ($N_F = 729$) ZigBee TNG feature set for $SNR = [0, 10, 18, 30]$ dB.

Figure IV-10 and Figure IV-11 consider the F-test p -values and test statistic values, respectively, through normalized histograms and the same bin widths as in Figure IV-8. Figure IV-10 and Figure IV-11 show a similar distributional issue for F-test p -values, where p -values are converging on 0 whereas the F-test statistic values do not converge to any one number.

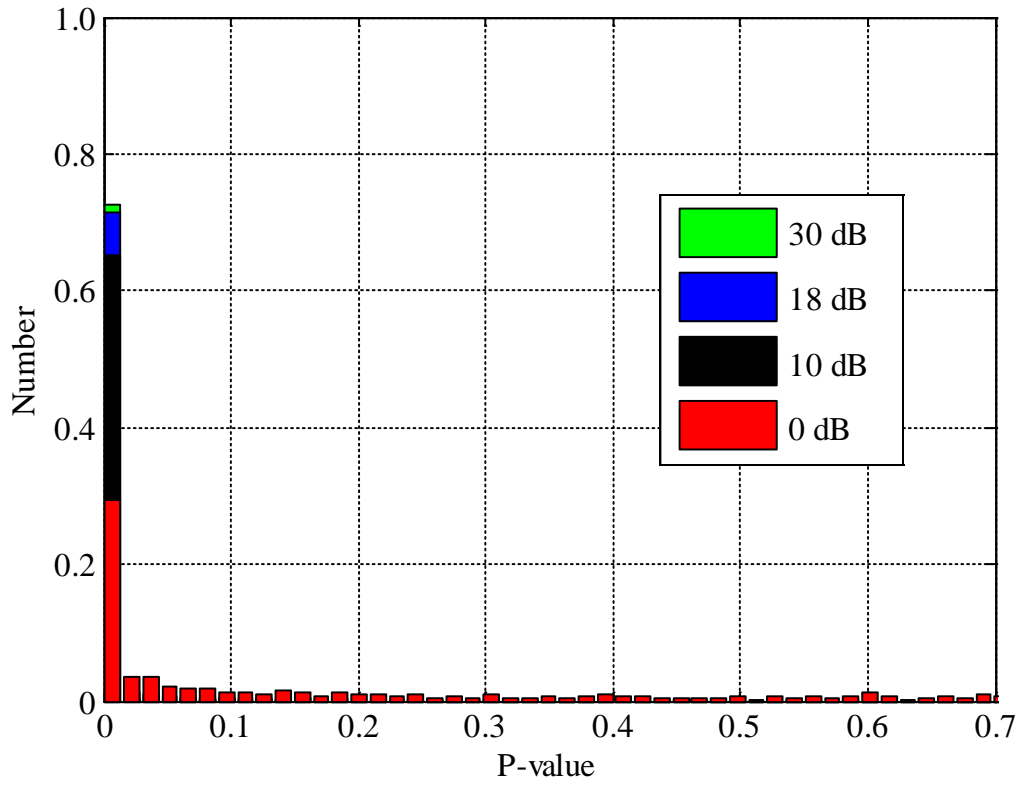


Figure IV-10: Normalized histogram of F-test p-values using a full-dimensional ($N_F = 729$) ZigBee TNG feature set for varying $SNR = [0, 10, 18, 30]$ dB.

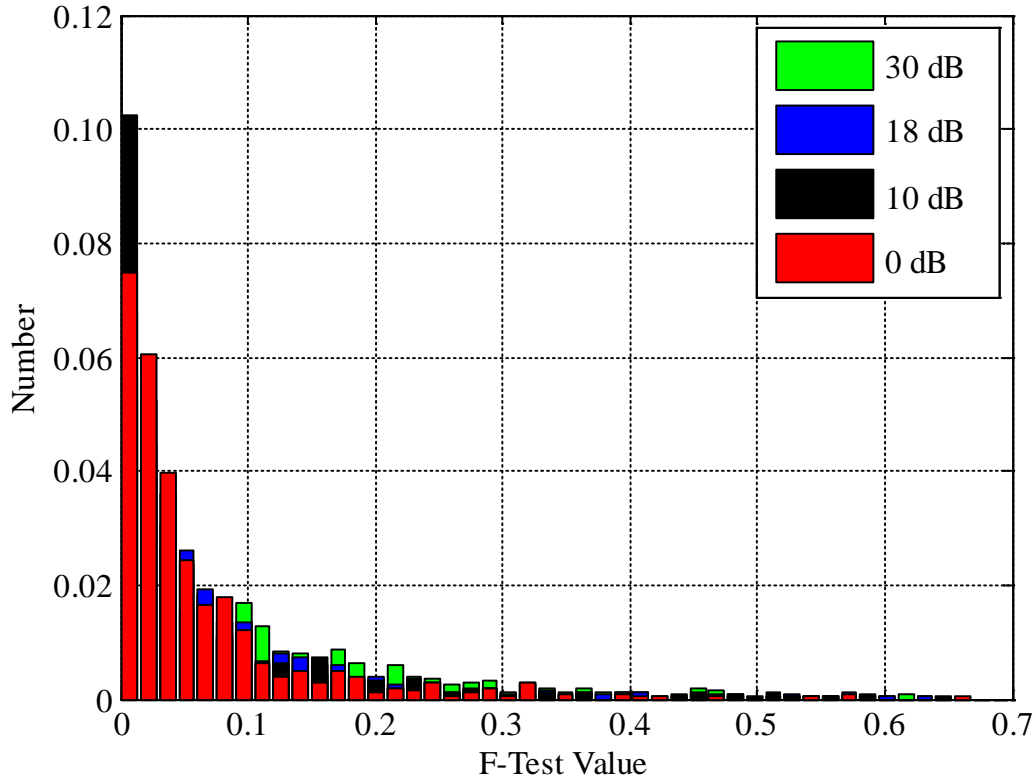


Figure IV-11: Normalized histograms of F -test statistic values using a full-dimensional ($N_F = 729$) ZigBee TNG features for $SNR = [0, 10, 18, 30]$ dB.

Table IV-1 condenses the results of Figure IV-8 and Figure IV-10 by illustrating that p -values trend towards 0, or indistinguishable numbers, as SNR increases. The general estimated decimal relative spacing between values of 2.22×10^{-16} , per [408], was used for this computation. Table IV-1 thus indicates that increasing signal strength corresponds to increasing significance. This result further mirrors that of p -values trending towards 0 in [365].

Table IV-1: Quantity of ZigBee p -values Less Than or Equal to 64-bit Relative Spacing, from [49].

METHOD	SNR			
	0 dB	10 dB	18 dB	30 dB
F-TEST P -VALUES	12	328	573	635
KS-TEST SUMMED P -VALUES	0	122	397	679

Table IV-2, adapted from Bihl et al. [49], further examines p -value and test statistic for ZigBee RF-DNA features the top 5 and bottom 2 ranked (by respective test statistic value) at $SNR = 10$ dB. Values in Table IV-2 are ranked by respective test statistic values for both F-test and KS-test, with the corresponding p -values. The 728th and 729th, lowest ranked values illustrate the scale of the values. While machine precision values are a continuum which rarely converge to any single number, noticeably many p -values are below the decimal relative spacing of 2.22×10^{-16} [408], and are thus notionally equivalent and equal to 0 for computing mean and variance. Evident in Table IV-2 is that ranking values equivalent to 0 may not provide a consistent means for ranking features and could be less effective when selecting a low number of features.

Table IV-2: P -values vs Test Statistic for ZigBee at SNR = 10 dB Ordered by Decreasing F-test and KS-Test Statistic Value, adapted from [49]

FEATURE	F-TEST		KS-TEST	
	TEST STATISTIC	P -VALUE	SUMMED TEST STATISTIC	SUMMED P -VALUE
1	542.64	$1.22 \cdot 10^{-303}$	3.316	$3.71 \cdot 10^{-94}$
2	471.78	$1.29 \cdot 10^{-268}$	3.251	0
3	432.97	$6.38 \cdot 10^{-249}$	3.242	$6.39 \cdot 10^{-97}$
4	424.26	$1.88 \cdot 10^{-244}$	3.169	$9.79 \cdot 10^{-98}$
5	420.74	$1.22 \cdot 10^{-242}$	3.053	$1.90 \cdot 10^{-61}$
⋮	⋮	⋮	⋮	⋮
728	0.280	0.839	0.164	2.18
729	0.043	0.988	0.150	2.67
VARIANCE	6,324.8	0.0094	0.2417	0.0646

Feature selection via p -values therefore has considerable issues. Further issues are illustrated in Appendix C where various academic datasets are considered through the KS-Test and F-test DRA methods. For both RF-DNA DRA and the academic datasets in Appendix C, test statistic values are seen to not converge on any specific number and thus they offer a more natural tool for feature comparison than p -values. Employing test statistic values for DRA is also consistent with the F-statistic DRA method formulated in [349]. As noted in Section 4.2.1.3(a), computing and interpreting p -values also involves further issues. Further comparisons of p -values versus test statistic values will be made via classification and verification performance.

4.2.2 Post-Classification Feature Selection DRA for RF-DNA

Model based feature selection methods involve computing a feature ranking as a byproduct or result of a classification model building process. Prior RF-DNA research has considered only GRLVQI feature relevance ranking and Random Forest as post-classification DRA methods. Although the MDA classifier has seen much use in RF-DNA applications, noticeably missing in previously applied DRA methods are MDA-based DRA methods. This absence is due to the assumption that MDA-based post-classification DRA was not directly possible [51, 91, 134]. However, various MDA-based DRA methods do exist in literature, e.g. [242, 351, 409], and these are further developed herein for application to RF-DNA. MDA based feature relevance ranking methods are considered and described below, including Wilk's Lambda, which examines the scatter matrices of MDA; Discriminant Weights, which are raw eigenvalues of the MDA matrices; and Discriminant Loadings, the correlation of the eigenvectors of MDA with the original data.

4.2.2.1 GRLVQI Feature Relevance Ranking

As discussed in Section III, GRLVQI feature relevance scores, ψ , provide a model-based indication of feature contribution to GRLVQI classifier development process [244–246, 266]. Prior work [89, 113] demonstrated ψ values offering comparable performance to KS-test p -value ranking for ZigBee feature selection with multiple discriminant analysis (MDA). Figure IV-12 examines GRLVQI relevance

scores, ψ , plotted by feature index number. Consistent with the KS-test and F-test DRA methods, GRLVQI relevance scores again show phase features as the most relevant.

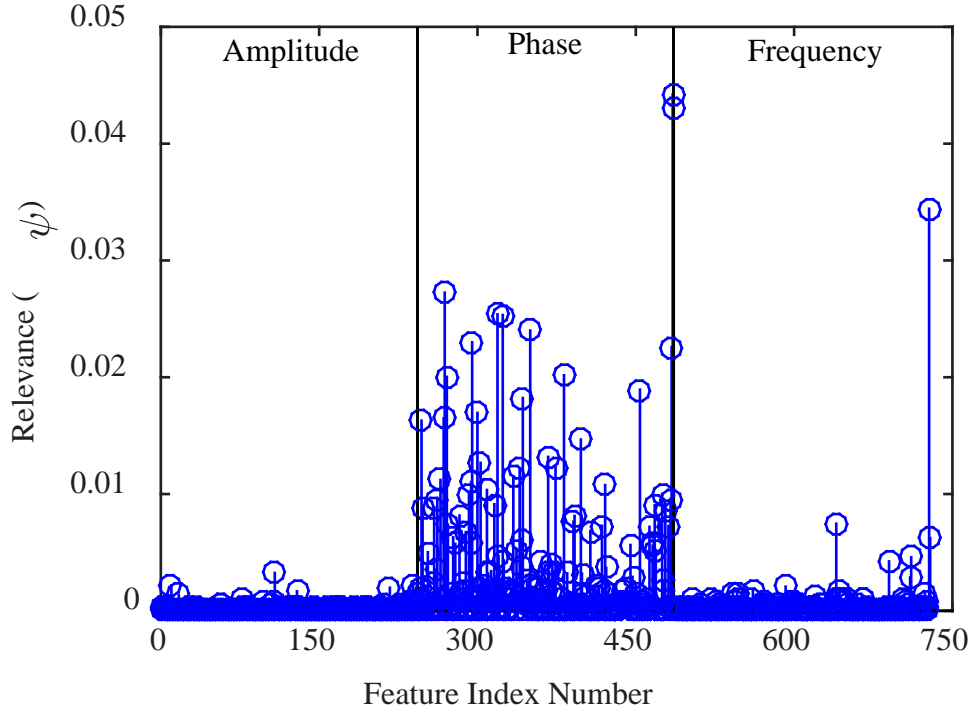


Figure IV-12: Feature ranking using GRLVQI relevance values using full-dimensional $N_F = 729$ ZigBee TNG observations at $SNR = 10$ dB.

4.2.2.2 MDA Based Feature Selection

Various methods of feature relevance ranking are implicit in MDA and can be determined relatively simply. Primarily, these methods involve ratios between scatter matrices and examining the discriminant functions themselves. Three general methods for MDA post-classification DRA will be considered: Wilk's Lambda, Discriminant Weights, and Discriminant Loadings.

(a) **Wilk's Lambda**

Wilk's Lambda values are computed via a ratio of determinants of MDA scatter matrices [409]; therefore this method is considered to be a post-classification DRA method. Wilk's Lambda has been used in various MDA application, e.g. [410, 411], and is computed as

$$\Lambda = \frac{\det \mathbf{S}_W}{\det \mathbf{S}_T}, \quad (4.9)$$

which is a ratio between the determinant of the within and total scatter matrices with $\Lambda \in [0 \ 1]$ [409]. In operation, large values of Λ indicate poor separation between groups, while smaller values of Λ indicate good separation between groups [409]. Logically, large group separations lend themselves to improved discrimination; therefore with lower Λ values are associated with more relevant features for classification [409].

The Wilk's Lambda method is used for DRA by computing each feature's Λ values using (4.9). For consistency with other DRA methods, herein Wilk's Lambda results are considered as $1 - \Lambda$, to ensure that higher values indicate more relevant features Figure IV-13 presents the $1 - \Lambda$ values for $SNR = 10$ dB for ZigBee. Consistent with the KS-test, F-test, and GRLVQI feature relevance ranking, the phase features appear most relevant in Figure IV-13.

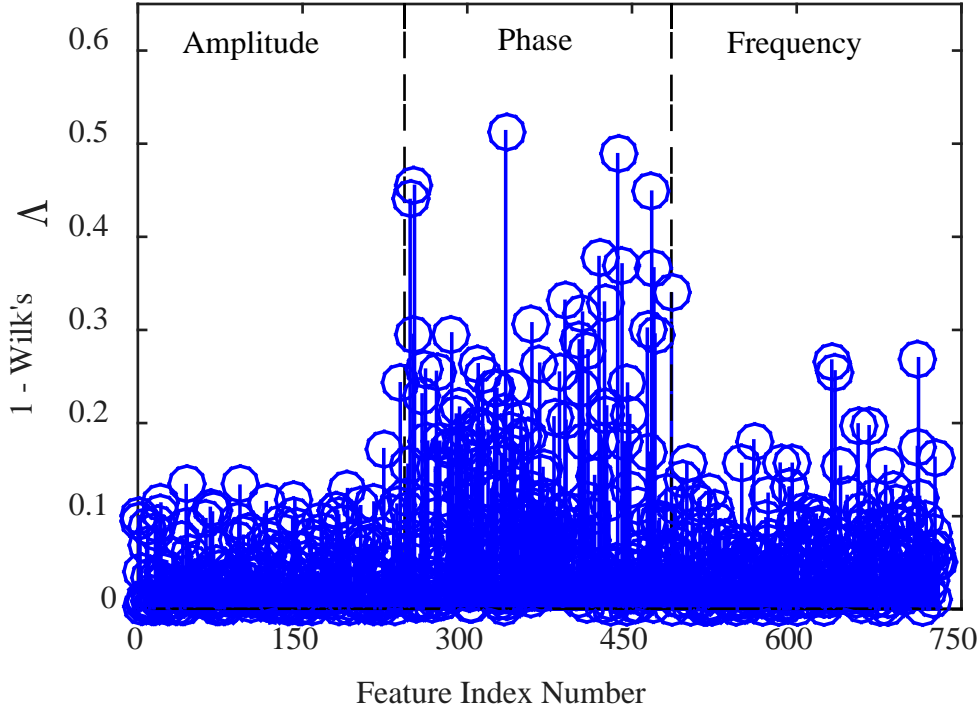


Figure IV-13: Feature ranking values from Wilk's Lambda ratio using full-dimensional $N_F = 729$ ZigBee TNG observation at $SNR = 10$ dB.

(b) Discriminant Weights and Group Means

One potential MDA-based DRA approach would be to remove features associated with relatively low eigenvector, or discriminant function coefficients, as employed in [412–414]. However, eigenvectors are considered to be generally unsuitable for providing feature relevance information [237], and this is considered imprecise for this purpose with small values can appear insignificant while actually being significant from an MDA standpoint [351]. For this reason, discriminant weights themselves are not considered for DRA. However, the basis of this approach, determining the connections

between discriminant functions and the data features, is similar to the discriminant loadings methods.

(c) **Discriminant Loadings**

Discriminant loadings were presented in Section 3.1.1, and are analogous to principal component loadings in describing how each feature contributes to a given projection vector [237, 415]. Visually examining MDA loadings is one approach to interpretation [416]. Figure IV-14 presents discriminant loadings for the $N_F = 729$ and $N_C = 4$ full-dimensional ZigBee TNG fingerprint set with values from (12) for $N_{Dim} = 3$ loadings vectors, as determined via (3.10). In Figure IV-14 both positive and negative MDA loadings values are visible. Also visible is an almost periodic sign change, which is possibly due to the binning process where adjacent bins could naturally be expected to have a directionally opposite action [417]. Also of interest is that the phase features appear to have higher magnitude loading values than amplitude and frequency, which is consistent with other DRA methods.

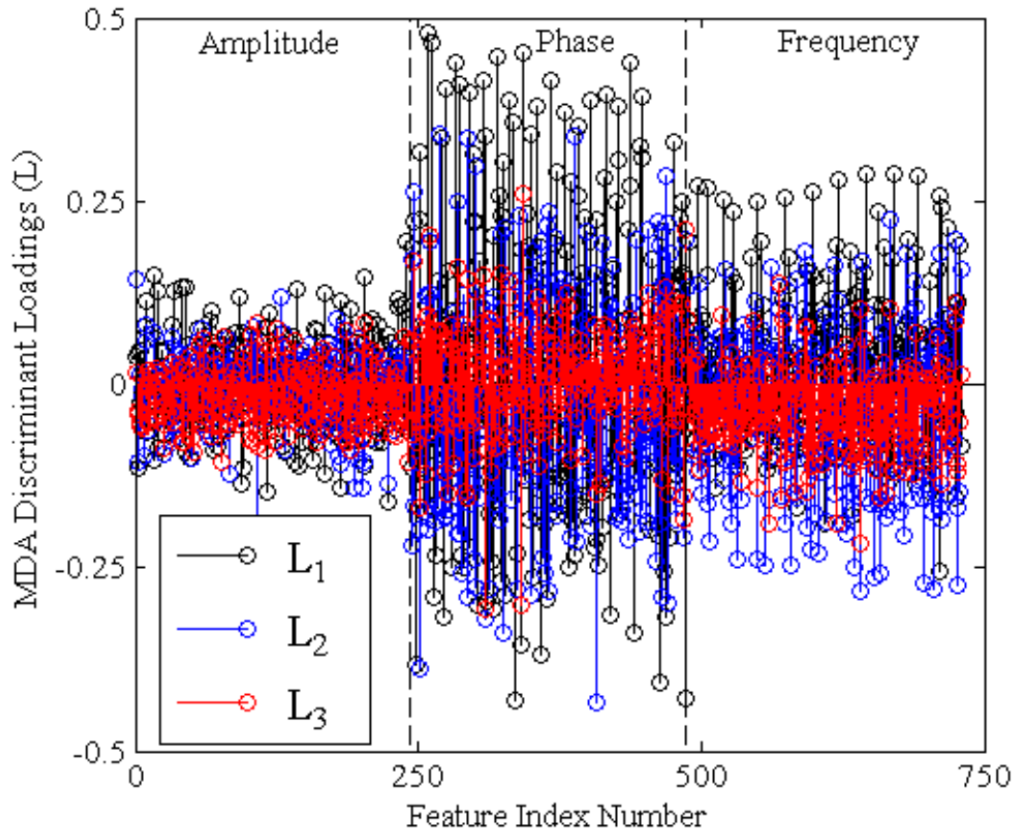


Figure IV-14: ZigBee discriminant loadings (L) for the three discriminant functions using full-dimensional $N_F = 729$ ZigBee TNG observations at $SNR = 10$ dB. Reprinted from [135].

However, apparent in Figure IV-14 is that each discriminant function presents different loading values for each fingerprint feature. Necessary in DRA is ranking each fingerprint feature with a single value and it is not readily apparent how to rank multiple loadings values for each feature. Therefore algorithmic fusion methods will be considered to develop an MDA loadings ranking method.

4.2.3 Algorithmic Fusion Methods

With multiple competing DRA methods used for feature selection, the combination of methods could be of interest. Fusion, in the signal processing sense, involves the combination of data, data features, or decisions from data for a combined result [418]. Fusion extends from Aristophanes' concept of *Φροντιστήριο*, or *phrontisterion*, the 'think tank' [419, p. 162; 420]. Of interest herein are 'fusing' various feature selection algorithms in an attempt to gain confidence in the features that are retained. To pursue this aim, a general review on fusion is needed. Figure IV-15 presents the three general types of fusion: data, feature, and decision. In general:

1. Data Level Fusion – combines the data from different sources; examples include combining a hyperspectral image pixel vector with the corresponding SAR intensity of that point [421] and combining different medical test values (e.g. blood sugar, enzymes, and etc.)
2. Feature Level Fusion – combines the extracted features in some manner to be input to a classifier/detector/etc., a few examples would include examining PCA vectors from two different data sources in an ANN as ANN inputs, or the addition of the patients address to the medical test values (in the above example)
3. Decision Level Fusion – combines the decision of multiple processes to create a combined decision. A few examples of this would be 1) applying multiple statistical classifiers to the same problem and then combining their result to create a final score, 2) including multiple doctors in a

patient's diagnosis, 3) combining a human interpretation of data with a computer decision (which might also be a fusion of multiple statistical classifiers too).

Additionally, variants on the architectures presented in Figure IV-15 can exist; for instance, Zhao et al. [422] created a combined feature-decision fusion approach with different feature subsets used for each classifier. The architecture of Zhao et al. [422] is therefore also a form of series fusion. Generally, either diversity and/or accuracy are used as measures for combining classifiers [423]. Recent results have indicated that classification consistently outperforms diversity when combining classifiers [423].

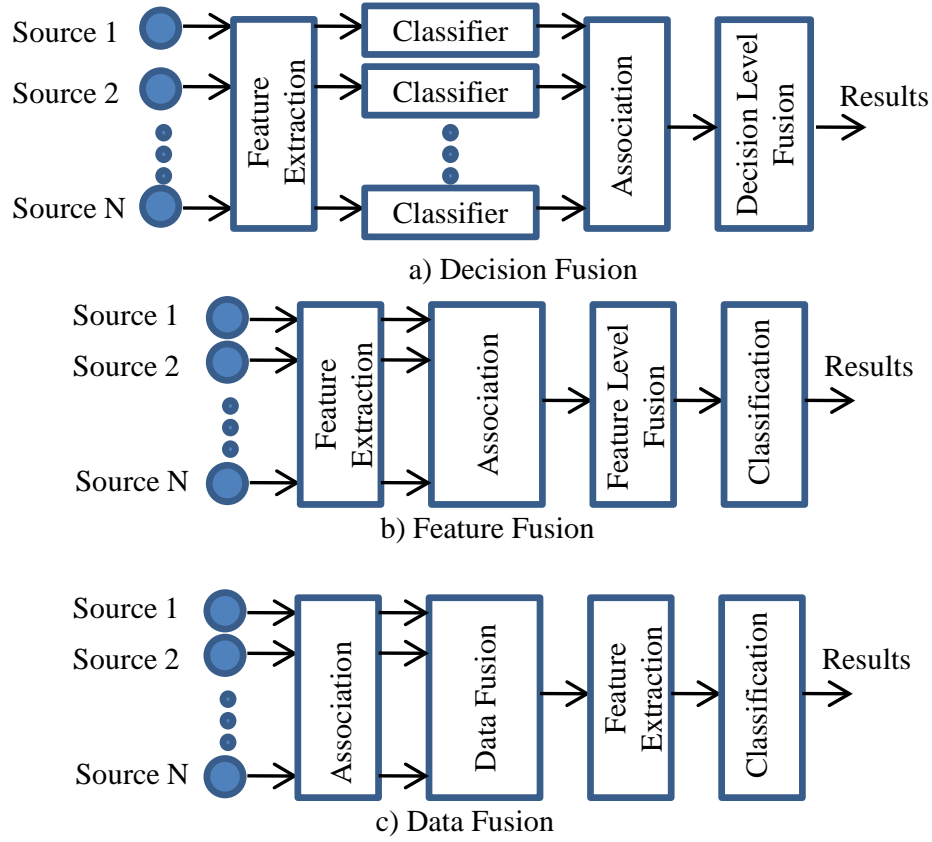


Figure IV-15: Three General Fusion Method Architectures, adapted from [418].

4.2.3.1 MDA Loadings Fusion (MLF)

As apparent in Figure IV-14 interpretation of MDA loadings into actionable feature rankings is non-trivial. Perreault et al. [424] introduced a composite *Potency* index,

$$\mathbf{L}_{Pot} = \mathbf{L}^2 \left[\frac{\lambda}{\sum_{i=1}^{N_{Dim}} \lambda_i} \right], \quad (4.10)$$

which both squares each loading value to remove interpretation issues associated with the direction of the loading considered combines and scales each loading value by the eigenvalue. Conceptually, the *Potency* index is a form of MDA Loadings Fusion (MLF), where loadings are fused through various methods to compute a final score. Although the *Potency* index has seen use in various MDA-based DRA application, e.g. [425–432], variations of this concept have not been explored. The *Potency* index and MLF methods are also conceptually similar to the weighted principal component approach of [433]; however, Kim and Rattakorn [433] considered variance explained and employed a moving range for selecting an appropriate level of dimensionality.

The following MLF strategies are therefore considered: first, *unscaled MLF*, where each loading for each feature will be considered as having an equal vote, second, *scaled MLF*, where each loading will be scaled by its relative weight as determined by the eigenvectors.

(a) **Unscaled MLF**

Thus, the following methodology was developed to create a single score for each fingerprint feature:

1. Compute the absolute value of all loadings vectors
2. Apply a fusion method (maximum or sum) to create a single vector for ranking features.

Two fusion methods were considered for Step 2, including 1) an *Unscaled Maximum* (*UMax*) score representing the *maximum* loading for each feature and 2) an *Unscaled*

Sum (*USum*) score representing the summation of loading values for each feature. The *USum* score is computed by summing the loadings, \mathbf{L} , across the columns, for the i^{th} feature this is computed as

$$\mathbf{L}_{USum,i} = \sum_{j=1}^{N_{DoF}} \mathbf{L}_i. \quad (4.11)$$

Similarly, the *UMax* score is computed by finding the maximum value of the loadings, \mathbf{L} , across the columns, for the i^{th} feature this is computed as

$$\mathbf{L}_{SSum,i} = \max(\mathbf{L}_i). \quad (4.12)$$

Results presented in Figure IV-16 display the *UMax* MDA loadings scores which show that phase features are again the most relevant for classifier model development.

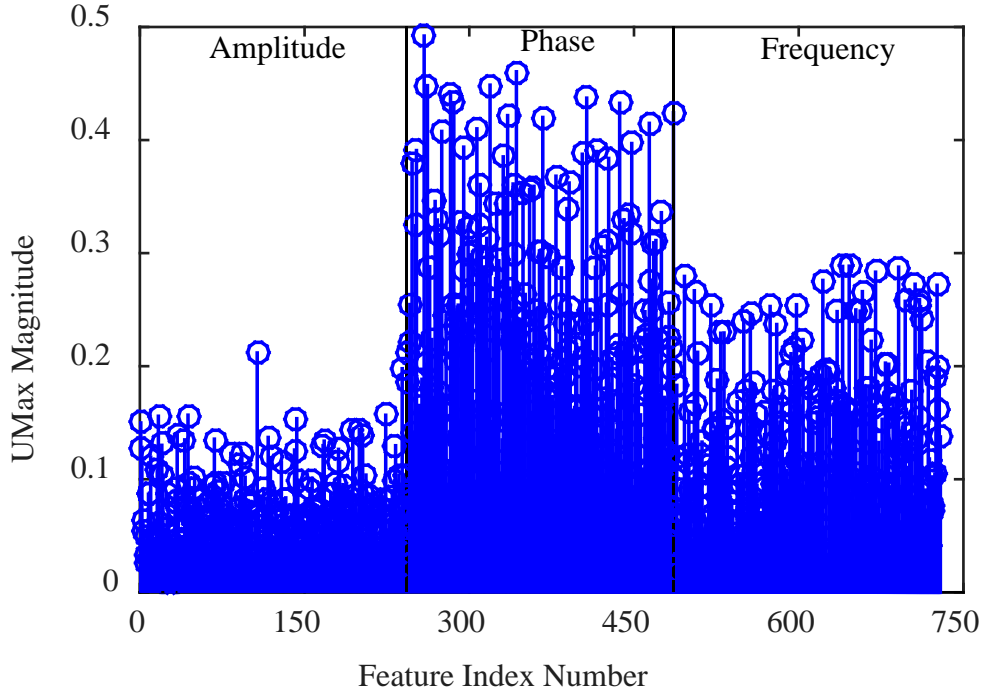


Figure IV-16: Feature ranking values from Unscaled Maximum ($UMax$) discriminant loadings using full-dimensional $N_F = 729$ ZigBee TNG observations at $SNR = 10$ dB.

(b) Scaled MLF

While the scaled MDA loadings presented in Figure IV-17 reflect overall how each feature is correlated to a given discriminant function, it ignores additional information contained in the Eigenvalues. Therefore a further MLF method, involving scaling the MDA loadings by their respective Eigenvalues, is a logical extension to account for the contribution that each discriminant function gives to total variance.

The loadings signify how each data feature is correlated to a given discriminant function. Because discriminant functions are also weighted by eigenvalue, it is not directly intuitive how to use them for feature selection. The method proposed involves

averaging the discriminant loadings after scaling them by their eigenvalue's contribution to total variance explained. This is computed as

$$\mathbf{L}_S = |\mathbf{L}| \left[\frac{\lambda}{\sum_{i=1}^{N_{Dim}} \lambda_i} \right], \quad (4.13)$$

which is very similar to the *Potency* index of [424] and (4.10), but avoids the squared loadings of (4.10) which shrink the overall MDA loadings magnitude.

This method enables the discriminant loadings to be ranked by the eigenvalue of each discriminant function and by the contribution of each feature to each discriminant function.

The following general methodology was used for Scaled MLF and is further described in [417]:

1. Compute the absolute value of all loadings vectors,
2. Multiply each absolute value loadings vector by the appropriate Eigenvalue-based weight per (4.13),
3. Apply a fusion method (maximum or sum) to create one vector for ranking features.

Consistent with *Unscaled MLF* are two fusion methods for Step 3: 1) a *Scaled Maximum (SMax)* score, and 2) a *Scaled Sum (SSum)* score. The *SSum* score is computed by summing the scaled loadings, \mathbf{L}_S , across the columns, for the i^{th} feature this is computed as

$$\mathbf{L}_{SSum,i} = \sum_{j=1}^{N_{DoF}} \mathbf{L}_{S,i}. \quad (4.14)$$

Similarly, the $SMax$ score is computed by finding the maximum value of the scaled loadings, \mathbf{L}_S , across the columns, for the i^{th} feature this is computed as

$$\mathbf{L}_{SSum,i} = \max(\mathbf{L}_{S,i}). \quad (4.15)$$

Figure IV-17 presents a series of scatterplots to show the general relationship between $UMax$, $USum$, $SMax$, and $SSum$ for the full-dimensional $N_F = 729$ feature set at $SNR = 10$ dB. As presented in [417], Figure IV-17 shows that the four fusion methods appear to largely provide different results with two exceptions: 1) that $UMax$ and $USum$ are correlated, and 2) that $SMax$ and $SSum$ are highly correlated. However, all four methods are further considered since small differences between methods can result in different DRA subsets and thus different results.

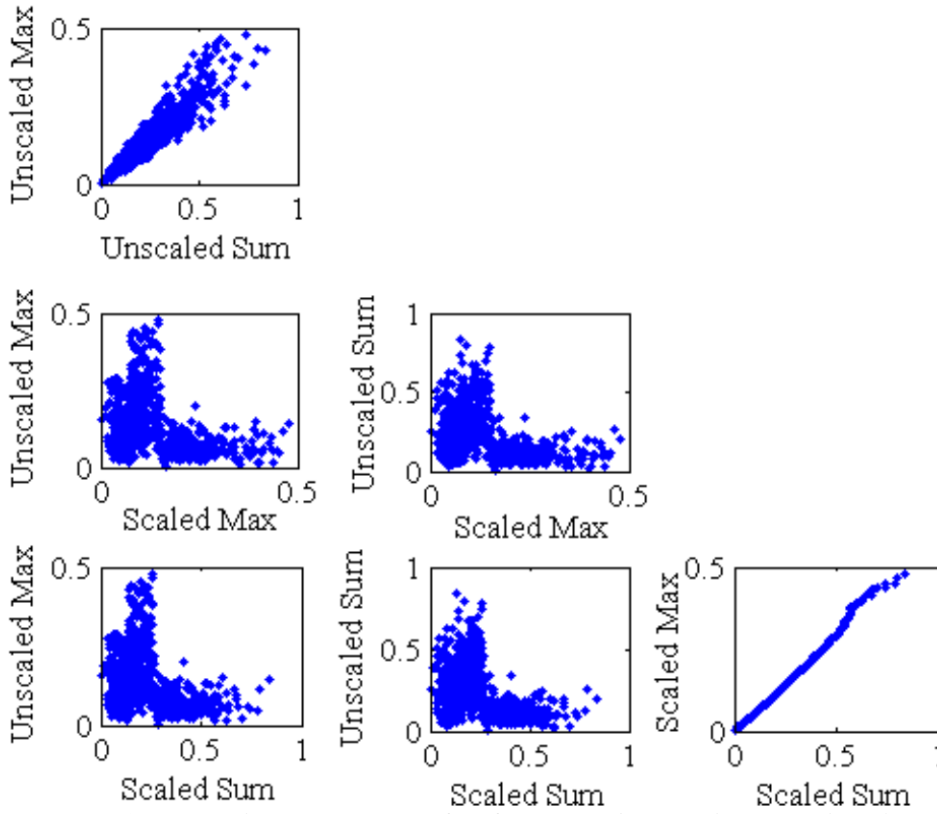


Figure IV-17: Matrix scatterplots for four MDA Loadings Fusion (MLF) methods, Unscaled (*UMax* and *USum*) and Scaled (*SMax* and *SSum*), using full-dimensional $N_F = 729$ feature set at SNR = 10 dB. Reprinted from [135].

4.2.4 DRA Fusion Methods

Herein, post-classification feature extraction, termed “DRA fusion,” is considered as an extension of decision fusion. Three DRA fusion methods are developed: rank-based DRA fusion, score-based DRA fusion, and concatenation DRA fusion.

4.2.4.1 Rank and Score Based Fusion

Rank and score based fusion extend series fusion by considering the DRA ranking scores for each feature. Both methods operate similarly and are conceptualized in Figure

IV-18. Step 1 in Figure IV-18 considers the ranks or normalized scores for each method, in Step 2 these are fused via summation and a new feature relevance ranking vector is computed.

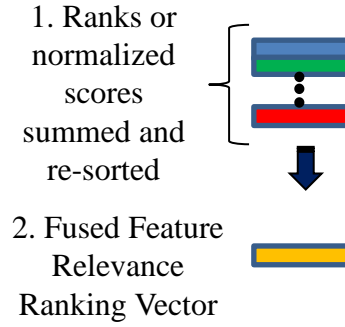


Figure IV-18: Generic Example of Score and Rank Fusion

(a) **Score Based DRA Fusion**

Score-based DRA, first normalizes the disparate DRA feature selection scales to a common scale via min-max data normalization,

$$\hat{X}_{min-max} = \frac{X - X_{min}}{X_{max} - X_{min}}, \quad (4.16)$$

where X is the original data, $\hat{X}_{min-max}$ is the scaled data, X_{min} is the minimum value, and X_{max} is the maximum value, can be used to place values on a $[0, 1]$ interval [434]. Although min-max normalization is sensitive to outliers [434], it is both a very common approach and places scores on an advantageous $[0,1]$ interval. Following normalization, scores from DRA methods are summed and then a new feature relevance ranking vector is computed from the fused scored.

(b) **Rank Based DRA Fusion**

Dichotomization involves converting a continuous variable into a discrete variable. An example of doing so would be converting continuous relevance scores into a ranked list, as described by [435]. Rank-based DRA fusion first considers the ordered ranking of each DRA method under consideration, these ranks are summed and a resulting summed rank vector is computed. The ordered rank of the summed rank vector is then used to determine feature relevance ranking. Thus rank-based DRA fusion is similar to score-based DRA fusion with the exception that the raw scores are not considered.

However, employing ranks may not be advantageous due to dichotomization issues. It is generally recommended to use continuous data, when available, rather than categorical data [436–441]. However, one encounters ranked lists in various feature relevance ranking operations and for RF-DNA rank-based DRA fusion avoids issues with score normalization, therefore considering the possibility of fusing results based on rank is considered.

4.2.4.2 Concatenation Fusion

Rank and score feature relevance ranking fusion seek to fuse the overall score of multiple feature relevance ranking methods. Concatenation fusion involves concatenating two or more vectors to form a single vector and has seen application in a variety of fields, c.f. [442–456]. Herein, an approach similar to that of Kekre et al. [457] is developed, where the selected features are appended to each other. However, care

must be taken in this process as multiple identical features will at a minimum add redundant features and necessarily introduce multicollinearity problems. Multicollinearity issues violate assumptions of MDA and other linear classifiers, therefore adding unique features is obvious necessary in feature selection fusion. Such a problem was not a concern for Kekre et al. [457] since they were fusing Red, Green, and Blue pixel information and hence was not concerned with uniqueness.

The RF-DNA concatenation DRA fusion method is conceptualized in Figure IV-19. Here, a user selects the desired total N_{DRA} and the $N_{DRA / method}$ top ranked features are proportionally taken from each DRA method,

$$N_{DRA / method} = \text{round}\left(\frac{N_{DRA}}{N_{methods}}\right), \quad (4.17)$$

where $N_{methods}$ are the number of DRA methods to be fused. The process in Figure IV-19 then removes repeated features to avoid singularity issues. The process then adds one next highest ranked feature from each DRA method and iterates until the fused vector has N_{DRA} features.

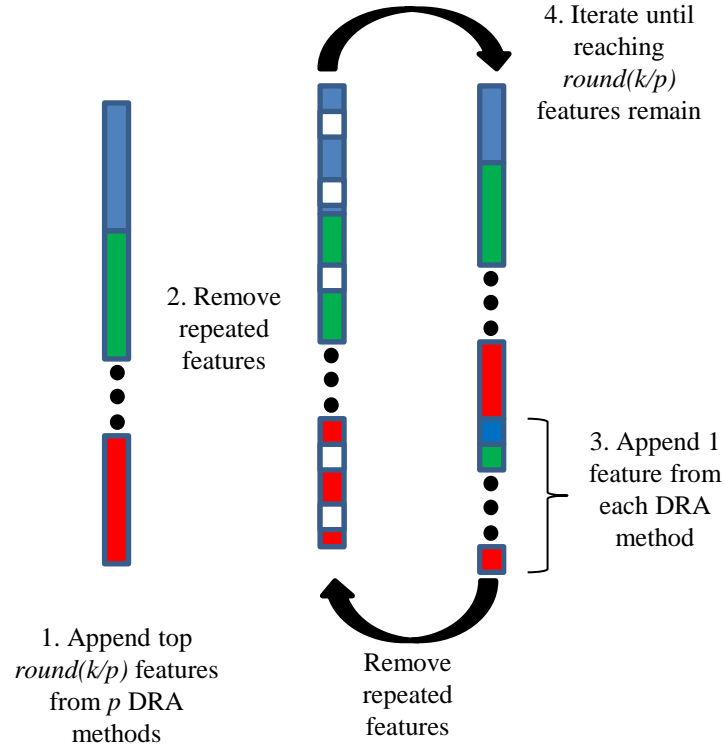


Figure IV-19: General Process for Concatenation Fusion

4.2.5 Random Feature Selection

When considering RF-DNA data, where there are hundreds of features, one could logically posit that any randomly selected and sufficiently large set of features could perform adequately. Since the ZigBee and Z-Wave RF-DNA datasets have no known corrupt features, it is very logical to believe that any random subset of features would offer some discriminating ability.

To account for this possibility, a random feature selection approach is considered to provide a lower bound for performance. For ZigBee, the random feature selection approach considers a uniform random feature relevance ranking values $U(0,1)$ for $N_F = 729$ feature set. An implicit assumption that higher magnitude random ranking values

are more relevant was used to select N_{DRA} feature sets. Since one random set of rankings may produce good results, replications are used and then classification and verification accuracies are averaged for the replicates. Performance from the random feature selection therefore offers a minimum expected level of performance for a given N_{DRA} .

4.2.6 Dimensionality Assessment

With relevance ranked features, DRA next involves selecting an appropriate level of dimensionality. Both qualitative and quantitative DRA dimensionality assessment methods are possible. Prior RF-DNA DRA research, e.g. [89, 113, 121], examined qualitative DRA for RF-DNA fingerprint features; however these were based on subjective assessments which may not be precise. Herein quantitative DRA approaches to estimate the intrinsic dimensionality in the data are developed. As noted by Jain et al. [213], an optimal approach to selecting features is via exhaustively examining classifier results produced from all possible combinations of features. However, this is very computationally intensive (and was noted as such by Jain et al. [213]) and is not practical for large datasets such as the ZigBee RF-DNA data where $N_{Feats} = 729$. Therefore quantitatively DRA approaches that examine intrinsic dimensionality of the data are developed and considered.

4.2.6.1 Qualitative Dimensionality Assessment

Prior RF-DNA work, c.f. [89, 113, 121] examined qualitative DRA methods for RF-DNA where subjective operator experience was used to select N_{DRA} . This was partially due to having no explicit selection criteria for selecting N_{DRA} based on KS-Test

p -value or GRLVQI relevance values. To determine an appropriate number of ranked features to retain, Dubendorfer et al. [113] examined various qualitative operating points corresponding to

$$N_{DRA} = [25, 50, 100, 200, 243] \quad (4.18)$$

feature sets. These were evaluated using an MDA/ML classifier, with the conclusion that $N_{DRA}=50$ features (selected using either KS-test p -values or GRLVQI relevance values) offered sufficient classification performance. However, this quantity or proportion (50/729, or 6.86% of the available features) is not necessarily generalizable to other RF-DNA fingerprint datasets and applications. Additionally, it is not known how to systematically search for these quantities. Therefore creating quantitative approaches based on the data itself are of particular interest.

4.2.6.2 Quantitative Dimensionality Assessment

Various quantitative dimensionality selection methods exist based on data covariance and correlation matrix responses [458–461]. Additionally, heuristics exist based on p -value significance and MDA-loadings magnitudes [358]. Of interest are developing quantitative dimensionality assessment methods for RF-DNA applications through data covariance and correlation matrices, p -values, and MDA-loadings.

(a) Heuristic-based Approaches on Discriminant Loadings

Discriminant loading magnitudes can also be used to estimate an appropriate number of features to retain. Various publications, c.f. [462–464], suggested that discriminate loadings magnitudes greater than 0.30 indicate a feature is significant.

Given that these works did not address scaled loadings, the heuristic value of 0.30 was applied to *Unscaled Max* scores at $SNR = 10$ dB and yielded $N_{DRA} = 51$ as the number of loadings greater than 0.30 in each composite. Because $N_{DRA} = 51$ is equivalent to the $N_{DRA} = 50$ determined by [113], this lends credence to the qualitative method of [113] and thus only $N_{DRA} = 50$ will be further examined for consistency with prior work.

(b) P-value based Approaches

Another approach to DRA assessment involves electing N_{DRA} from p -value significance [358]. As described in Section (b) p -values tend to zero for RF-DNA fingerprints and thus employing a p -value threshold for quantitative DRA could involve retaining a majority of the data. For instance, at 10dB, if one employed a p -value threshold of 5%, a common statistical significance threshold, one would retain $N_{DRA} = 674$ if using the F-test or $N_{DRA} = 512$ if using the KS-test.

Table IV-3 further presents the quantity of retained features using the F-test and KS-test at $SNR = [0, 10, 18, 30]$ dB for different statistical significance levels. Statistical significance levels of [0.1%, 1%, 5%, 10%] are employed as commonly used [465], although largely arbitrary [379], statistical thresholds. Comparing Table IV-3 with the results of [121] indicates that p -value DRA assessment heavily over-estimates the number of features to retain since phase (ϕ) features, $N_F=243$ herein, are known to offer performance comparable to the baseline. Therefore, p -value dimensionality assessment appears neither appropriate or is considered for ZigBee RF-DNA data.

Table IV-3: Dimensionality Assessment by p -value and Significance Level, Reprinted from [49].

SNR	METHOD	SIGNIFICANCE LEVEL			
		<i>0.1%</i>	<i>1%</i>	<i>5%</i>	<i>10%</i>
0 dB	F-TEST P-VALUES	196	264	350	402
	KS-TEST SUMMED P-VALUES	37	74	130	160
10 dB	F-TEST P-VALUES	589	639	674	688
	KS-TEST SUMMED P-VALUES	337	414	512	557
18 dB	F-TEST P-VALUES	706	713	720	722
	KS-TEST SUMMED P-VALUES	666	692	711	716
30 dB	F-TEST P-VALUES	718	725	727	728
	KS-TEST SUMMED P-VALUES	727	729	729	729

(c) **Data Covariance Matrix Approaches**

DRA assessments on the intrinsic dimensionality in data can also be considered. If one considers the eigenvalues of the data covariance (or correlation matrix) one can estimate data dimensionality based. Given that RF-DNA features have consistent units, the covariance matrix was considered herein with three quantitative DRA assessment methods: Kaiser's Criterion, Maximum Distance Secant Line (MDSL), and Horn's Curve.

(i) **Kaiser Criterion**

Kaiser criterion offers a basic estimate of N_{DRA} with Eigenvalues greater than the average eigenvalue being retained [237, 458, 466]; when correlation eigenvalues are considered, this results in all eigenvalues greater than 1 being retained [467]. Although it can offer reasonable performance, it is also acknowledged as a rather arbitrary method

[458]. Because this metric is frequently generalized to just selecting the eigenvalues above 1, both the appropriate metric (above the mean) for covariance eigenvalues is presented along with the ‘above 1’ metric.

Kaiser criterion offers a basic estimate of dimensionality with the DRA assessment made where the quantity of covariance matrix eigenvalues greater than the mean are retained [237, 458]. Although offering reasonable performance, Kaiser is acknowledged as a rather arbitrary method [458]. Kaiser’s criterion at SNR = 10 dB suggests retaining $N_{DRA} = 191$ features.

(ii) **Cattell’s Scree Plot**

One extension of the Kaiser criterion involves including visual subjectivity in the form of Scree plots. Scree plots involve two dimensional plots of data covariance (or correlation) matrix Eigenvalues versus rank order, and provide a visual method of determining the dimensionality of the data [237]. Cattell’s Scree Test, involves visually examining the scree plot and selecting N_{DRA} above the inflection point, the proverbial ‘elbow in the curve’ [458]. The difficulty of this methods involves selecting the actual inflection point and N_{DRA} .

1. **Maximum Distance Secant Line (MDSL)**

The MDSL approach, introduced by Johnson et al. [468], aims to remove subjectivity from Cattell’s Scree Test through algorithmic means. MDSL both removes subjectivity of Cattell through automation, where 1) one creates a line between the first and last rank ordered eigenvalues and 2) on then finding the point with the largest

perpendicular distance from this line, i.e., the inflection point [468]. Using MDSL at $\text{SNR} = 10 \text{ dB}$ $N_{DRA} = 26$ features would be retained.

(iii) **Horn's Curve**

Horn's curve is another eigenvalue based DRA assessment method where eigenvalues are computed for a random dataset of the same size and rank as the ZigBee fingerprint set under analysis [469]. Horn's curve involves plotting the data sample correlation matrix eigenvalues against the Horn's curve eigenvalues [469]. The intrinsic dimensionality of the data is determined by counting the number of data eigenvalues that appear above Horn's curve [469]. Using the Horn's curve algorithm of Bigley [466], at $\text{SNR} = 10 \text{ dB}$ Horn's curve indicated $N_{DRA} = 157$ features should be retained.

4.2.6.3 DRA Assessments and ZigBee RF-DNA Features

As all of the presented DRA assessments provided different N_{DRA} subsets, multiple DRA subsets must be considered. For comparison with qualitative methods, $N_{DRA} = [50, 100]$ subsets are examined for consistency with [113], additionally a lower qualitative DRA assessment of $N_{DRA} = 10$ is also important to examine to understand performance when only a very limited subset of features are available and thus examine how DRA methods fundamentally interacts with classifier performance. The resultant N_{DRA} subsets to examine for competing DRA methods is thus

$$N_{DRA} = [10, 26, 50, 100, 157, 191] , \quad (4.19)$$

which considers both quantitative and qualitative methods. Comparison with the full-dimensional $N_{DRA} = 729$ feature set is also requisite to generate a performance baseline for comparison.

4.3 DRA Applications to ZigBee Data

To understand and compare the presented DRA methods, first a simple comparison of DRA methods results through correlation will be considered. Then a comparison of how different DRA methods select different features will be discussed. Finally, a comparison of classification and verification performance assessments, with the MDA/ML classifier, will be made using the ZigBee dataset.

4.3.1 DRA Method Comparisons

Consistency was seen in the KS-test, F-test, GRLVQI relevance values, and MDA loadings where phase (ϕ) features are noticeably more relevant than both Amplitude (a) and Frequency (f) features. This observation is further consistent with [89, 113], which concluded that Phase (ϕ) features alone are typically the most relevant for reliable device discrimination.

However, it's not apparent that each method scores the same features similarly. Table IV-4 presents a correlation matrix using Pearson correlations at $SNR = 10$ dB, where most methods are seen to be not highly correlated in their scores. Incidentally, both GRLVQI relevance and random loadings were the least correlated to any other method, indicating limited similarity to the other methods. $SSum$ and $SMax$ were highly

correlated, while the other loadings methods are less correlated, thus indicating that loadings methods are sensitive to the fusion method.

In Table IV-4, both the KS-test and the F-test are seen to be highly correlated, which indicates that both methods achieve similar results. This is largely a logical result because both methods are univariate, distribution based, and consider a given feature and a vector of categorical class identities. The F-test result was also highly correlated with *USum* and *UMax*, mirroring the results of [462] which reported a positive correlation of 0.675 between DRA loadings and the F-test.

Table IV-4: Correlation Matrix for DRA Method Scores at $SNR = 10$ dB, from Bihl et al. [135]. High correlations (>0.8) and low correlations (<0.2) are in bold and shaded light grey.

		DRA Feature Selection Method								
		<i>Pre-classification</i>			<i>Post-classification</i>			<i>Baseline</i>		
		KS	F-Test	GRLVQI	Wilk's	MLF <i>SMax</i>	MLF <i>SSum</i>	MLF <i>UMax</i>	MLF <i>USum</i>	Random
<i>Pre-classification</i>	KS	1.0	0.665	-0.164	0.388	0.474	0.166	0.726	0.6977	-0.038
	F-Test		1.0	-0.130	0.749	0.590	0.264	0.928	0.890	0.011
<i>Post-classification</i>	GRLVQI			1.0	-0.082	-0.094	-0.030	-0.167	-0.178	0.041
	Wilk's				1.0	0.377	0.144	0.730	0.726	-0.037
	MLF <i>SMax</i>					1.0	0.8589	0.630	0.565	-0.035
	MLF <i>SSum</i>						1.0	0.257	0.253	-0.046
	MLF <i>UMax</i>							1.0	0.937	-0.004
	MLF <i>USum</i>								1.0	-0.012
<i>Baseline</i>	Random									1.0

Since Table IV-4 illustrates that each DRA method is ranking features differently, examining the top ranked features across DRA methods is of interest. Consistent with [417], Figure IV-20 considers the top $N_{DRA} = 10$ features through a bar plot showing which features are selected for each method. Only one replicate of the Random Selection DRA method presented for brevity. Of interest in Figure IV-20 is that, although most features selected are Phase (ϕ) features (indices 244 to 486) most DRA methods selected entirely different features [417]. Interestingly, a few features in Figure IV-20 were consistently selected by multiple methods, thus indicating that some features are predominantly important, an observation consistent with results in [89, 113].

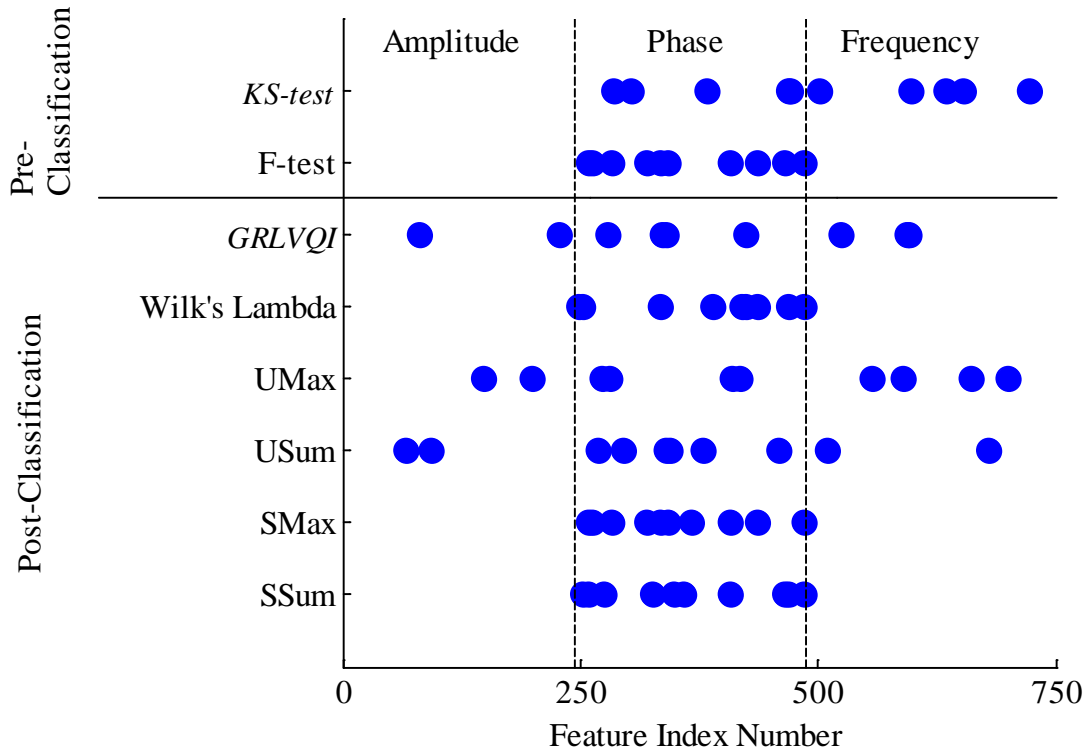


Figure IV-20: Top ranked $N_F = 10$ reduced dimensional feature sets by DRA method, reprinted from [135].

Figure IV-21 and Figure IV-22 further consider the differences in DRA method feature ranking for $N_F = 26$ and $N_F = 50$, respectively. While the figures are consistent with those of Figure IV-20, where methods largely select different features, as N_F increases, it is apparently that DRA methods begin to select similar features, which are predominantly phase features.

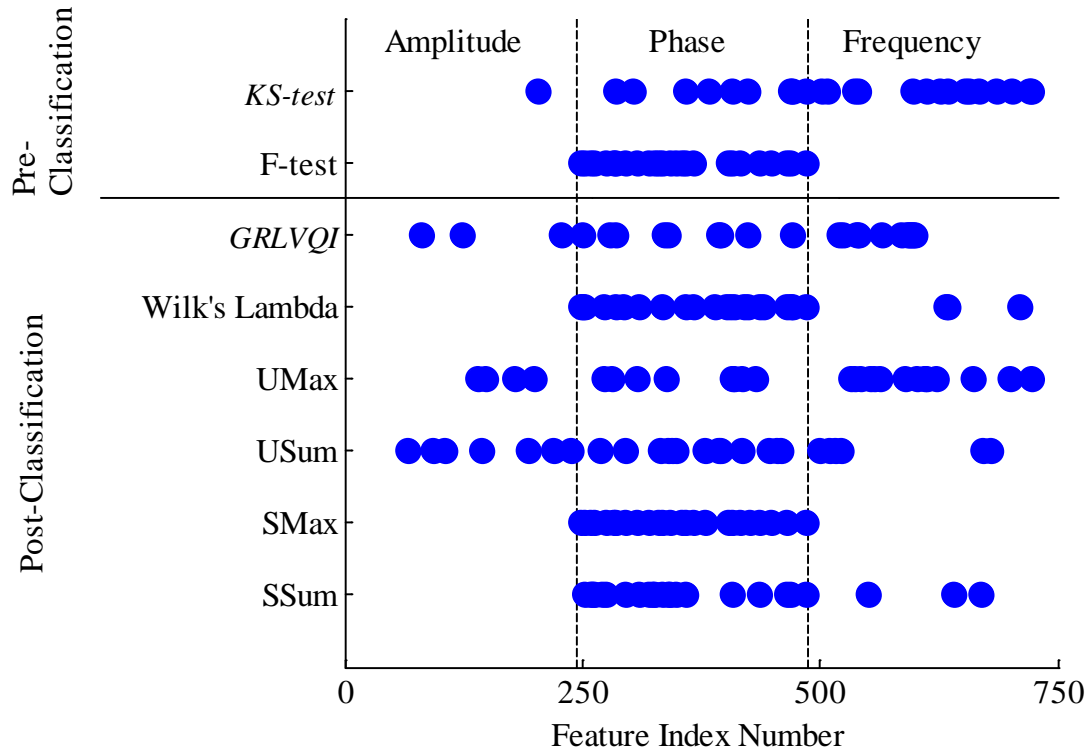


Figure IV-21: Top ranked $N_F = 26$ reduced dimensional feature sets by DRA method.

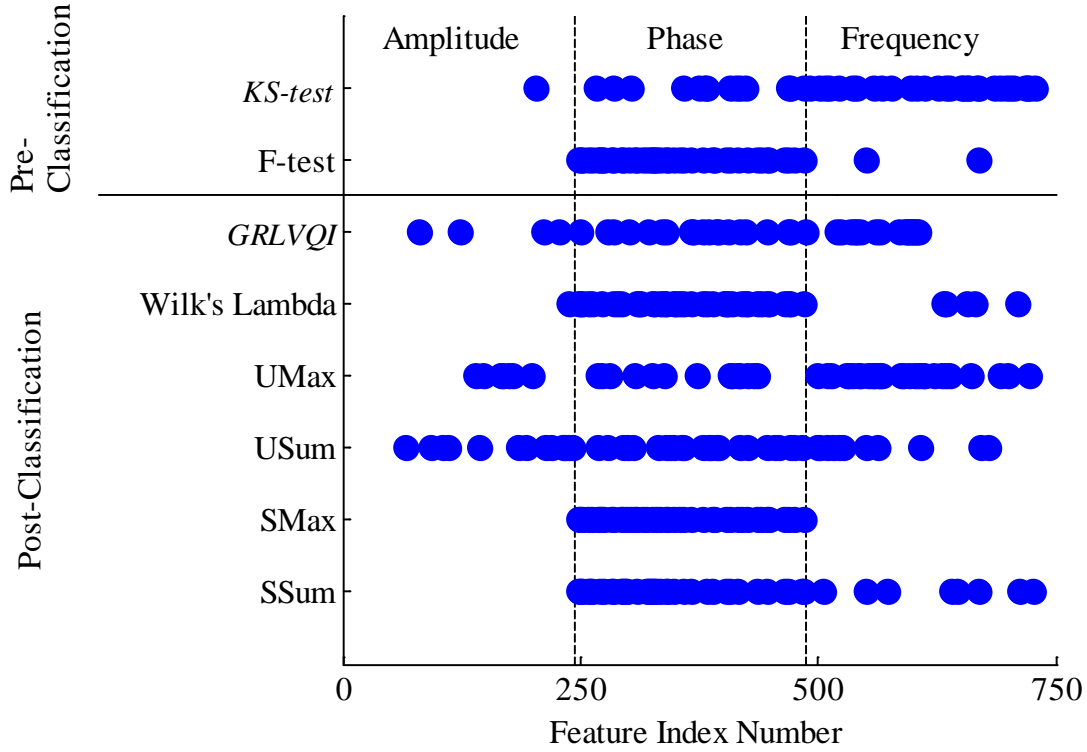


Figure IV-22: Top ranked $N_F = 50$ reduced dimensional feature sets by DRA method.

Table IV-5 further examines the features selected by each DRA method per each DRA subset. In Table IV-5, the collective total features selected, N_{TOT} , for F-test, KS-test, GRLVQI, Wilk's Lambda, $USum$, $UMax$, $SSum$, and $SMax$, are presented for each N_{DRA} subset. When considering $N_{DRA} = 10$, $N_{TOT} = 61$ total features were selected; however, 78.7% of these 61 features were uniquely selected by only one DRA method and hence many features were selected by multiple DRA algorithms. Table IV-5 presents additional information regarding the percentage of N_{TOT} which are amplitude (a), phase (ϕ), and frequency (f) features, and the percentage of variance (σ^2), skewness (γ), and kurtosis (κ) statistics. Notable throughout DRA subsets, and consistent with [89, 113], is

that majority of features selected are phase features. No obvious biases are seen toward variance (σ^2), skewness (γ) or kurtosis (κ) being selected.

Table IV-5: DRA Subset Statistics for F-test, KS-test, GRLVQI, Wilk's Lambda, $USum$, $UMax$, $SSum$, and $SMax$. Reprinted from [135].

DRA SUBSET	N_{Tot}	% UNIQUE	(a, ϕ , f) %	(σ^2 , γ , κ) %
NDRA = 10	61	78.7%	7.5, 73.8, 18.7	32.5, 46.3, 21.2
NDRA = 26	142	72.5%	7.2, 65.9, 26.9	34.6, 38.0, 27.4
NDRA = 50	238	65.1%	7.0, 64.3, 28.7	37.8, 35.2, 27.0
NDRA = 100	381	48.8%	7.1, 57.1, 35.8	38.6, 32.3, 29.1
NDRA = 157	505	39.2%	7.5, 54.9, 37.6	38.3, 31.7, 30.0
NDRA = 191	545	31.9%	8.1, 53.5, 38.4	37.7, 32.5, 29.8

4.3.2 DRA Method Classification Performance Assessments

Beyond comparing DRA methods statically, further comparison of DRA methods through MDA/ML classification accuracy on the ZigBee RF-DNA dataset need consideration. Representative MDA/ML average TST %C versus SNR results are presented in Figure IV-23 to Figure IV-25. Figure IV-23 presents results from the MDA/ML model using $N_{DRA} = 10$, Figure IV-24 presents results from the MDA/ML model using $N_{DRA} = 26$, and Figure IV-25 presents results from the MDA/ML model using $N_{DRA} = 50$. Additional results from $N_{DRA} = [100, 157, 191]$ are presented later in tables.

Although at $N_{DRA} = 10$ no feature selection method achieves the %C > 90% benchmark, and thus relative dB gain is not computed for comparison, the results here in Figure IV-23 show DRA performance differences across methods. Consistent with

[395], Figure IV-23 shows MLF-based methods as offering significantly higher performance than other DRA methods with MLF methods having a 10% improvement in $\%C$ for most of the SNR considered when compared to other methods. Additionally, MLF methods have an SNR gain over competing DRA methods of 10 to 12 dB for $60\% < \%C < 75\%$ (max). The results of $N_{DRA} = 10$ suggest that MLF-based DRA methods perform better than competing methods here since MLF feature relevance rankings were computed close to the functions used for MDA classifier development.

Results for $N_{DRA} = 10$ and $N_{DRA} = 26$, respectively Figure IV-24 and Figure IV-25, show that all feature selection methods tend to achieve similar performance as the number of features considered increases [417]. Despite this, some differences are still seen in the performance offered by the DRA methods with the loadings based-methods again offering significantly higher performance than the other methods under analysis.

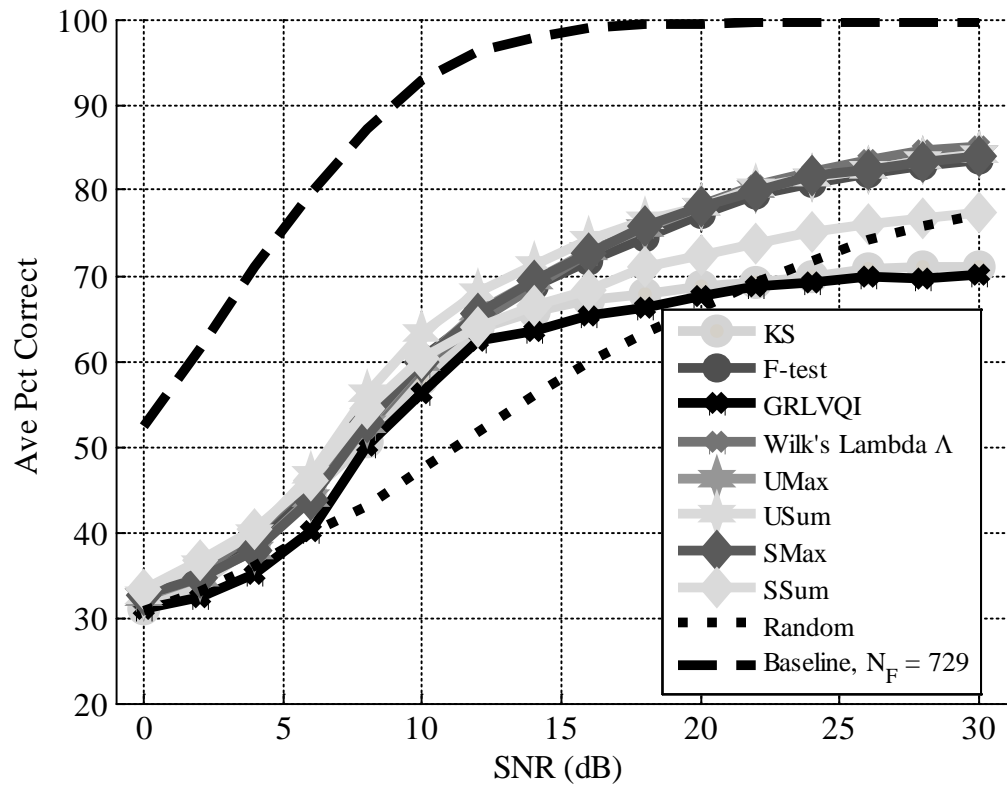


Figure IV-23: ZigBee MDA/ML Testing (TST) classification performance for $N_{DRA} = 10$ reduced dimensional feature sets, reprinted from [135].

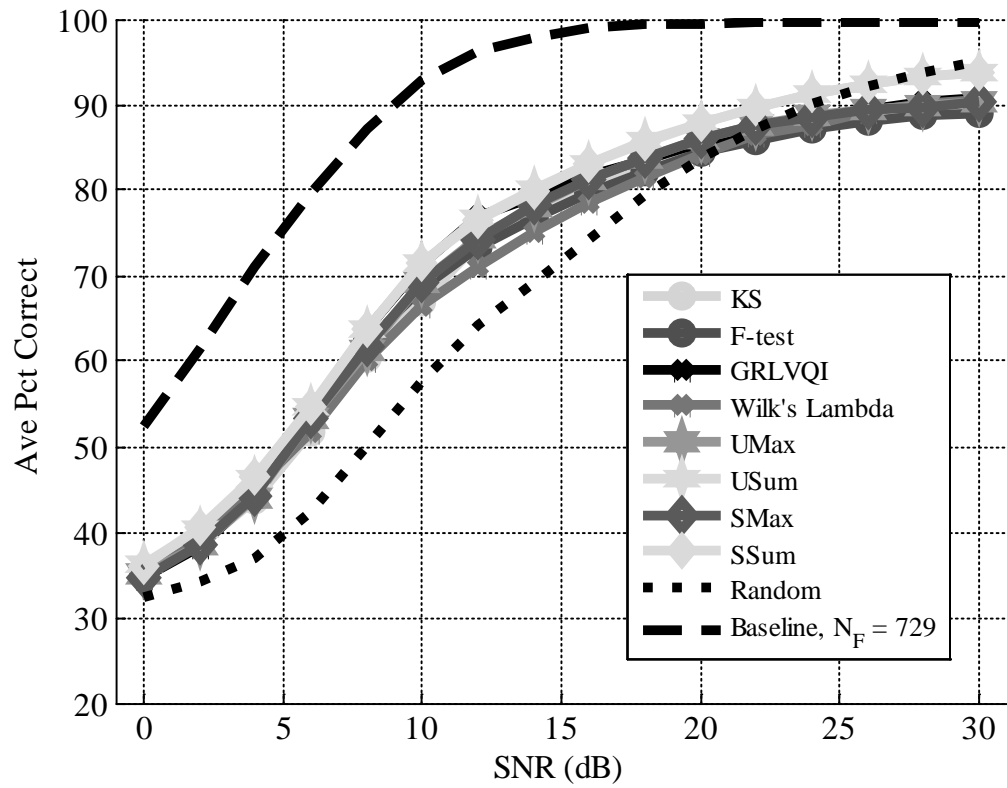


Figure IV-24: ZigBee MDA/ML *Testing (TST)* classification performance for $N_{DRA} = 26$ reduced dimensional feature sets, reprinted from [135].

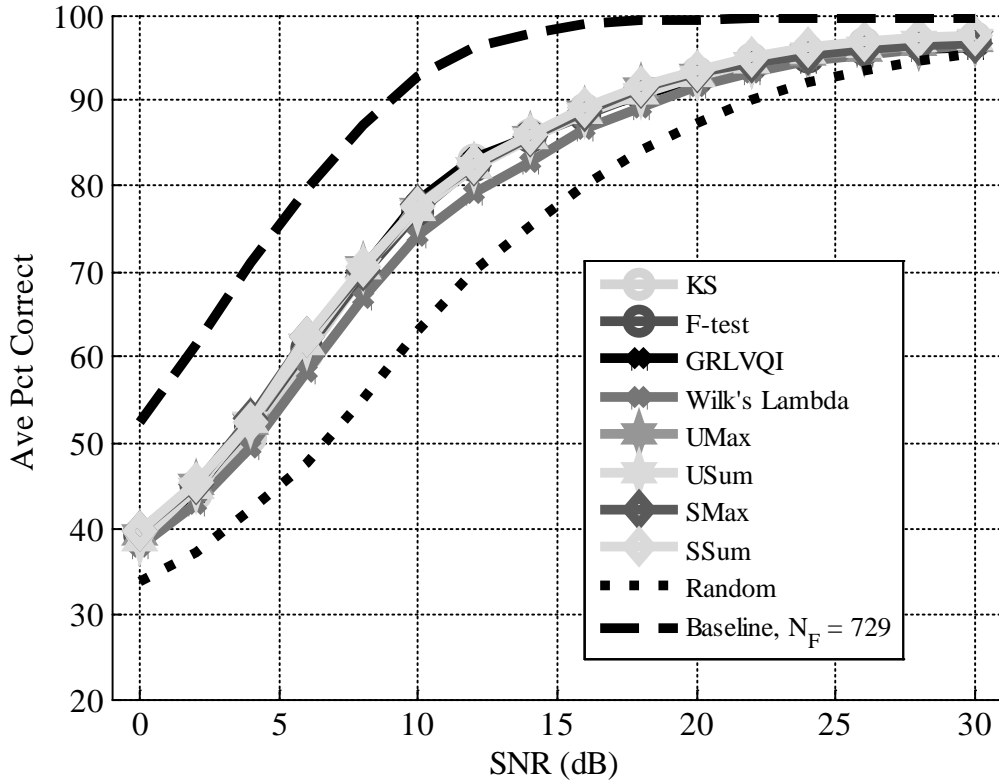


Figure IV-25: ZigBee MDA/ML Testing (TST) Classification performance for $N_{DRA} = 50$ reduced dimensional feature sets, reprinted from [135].

Figure IV-23 through Figure IV-25 represent only a few instances showing the relationship between N_{DRA} and classification performance. To further understand how DRA influences performance, Figure IV-26 considers classification performance and dimensionality of each DRA method at $SNR = 10$ dB. In Figure IV-26 additional N_{DRA} subsets, $N_{DRA} = [250, 300, 350, 400, 450, 500, 550, 600, 650, 700]$, are considered along with those of (4.14). Figure IV-26 shows an expected decrease in classification accuracy as one decreases N_{DRA} , which is especially seen for $N_{DRA} < 200$. Consistently high performance is further seen across all methods except Wilk's and Random.

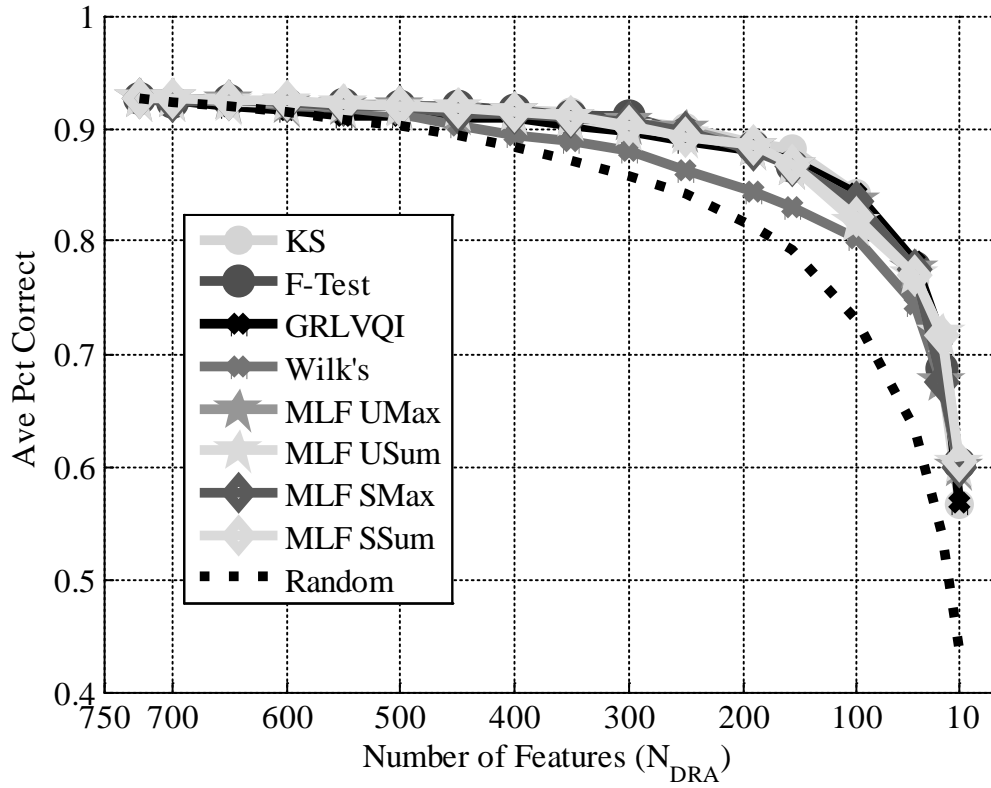


Figure IV-26: ZigBee MDA/ML Testing (TST) classification performance for each DRA method at SNR = 10 dB. $N_{DRA} = [10, 26, 50, 100, 157, 191, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700]$ reduced dimensional feature sets are evaluated to understand how DRA fundamentally impacts performance. Reprinted from [135].

Table IV-6 reproduces a table in [135] by presenting gain tradeoff values at $\%C > 90\%$ for all DRA methods and all N_{DRA} levels of dimensionality from (4.13). However, Table IV-6 presents no values for $N_{DRA} = 10$ since no DRA methods achieved $\%C > 90\%$ at this level of dimensionality [135]. Gain tradeoff values in Table IV-6 show a considerable advantage of MLF methods at $N_{DRA} = 26$ over other methods, where MLF methods achieve better performance than either GRLVQI or Wilk's; additionally, *SMax*,

SSum, and *UMax* achieve better performance than randomly selected sets. Incidentally, the MDA/ML model developed using either KS-test and F-test selected features do not achieve $\%C > 90\%$ at $N_F = 26$. As N_{DRA} increases to $N_{DRA} = 157$ and $N_{DRA} = 191$ it is seen that the competing DRA methods offer comparable classification performance.

Table IV-6: Relative DRA “Gain” (dB) Over Baseline Performance for %C = 90% Classification Accuracy. Bold entries with light grey shading denote best case (lowest gain) performance and bold entries denote values within 10% of the best. Reprinted from [135].

DRA SUBSET		DRA FEATURE SELECTION METHOD								
		<i>PRE-CLASSIFICATION</i>		<i>POST-CLASSIFICATION</i>						<i>BASELINE</i>
		KS	F-TEST	GRLVQI	WILK’S	MLF <i>S</i> _{MAX}	MLF <i>S</i> _{SUM}	MLF <i>U</i> _{MAX}	MLF <i>U</i> _{SUM}	RANDOM
$N_{DRA} = 26$	TNG	*	*	-18.747	-18.727	-14.269	-13.347	-13.809	-14.607	-14.937
	TST	*	*	-19.349	-19.967	-14.167	-13.817	-13.847	-14.967	-15.407
$N_{DRA} = 50$	TNG	-7.877	-8.337	-8.357	-9.617	-7.947	-7.697	-7.897	-9.957	-13.557
	TST	-8.077	-8.687	-8.787	-10.157	-8.347	-7.967	-8.387	-10.137	-13.007
$N_{DRA} = 100$	TNG	-4.707	-4.587	-3.387	-5.577	-4.137	-4.817	-4.127	-5.747	-8.997
	TST	-4.887	-4.817	-3.407	-5.987	-4.487	-4.957	-4.477	-6.067	-8.777
$N_{DRA} = 157$	TNG	-2.747	-2.627	-2.207	-4.287	-2.647	-2.487	-2.507	-2.727	-5.317
	TST	-2.927	-2.787	-2.357	-4.407	-2.937	-2.587	-2.727	-2.757	-4.957
$N_{DRA} = 191$	TNG	-2.007	-1.907	-1.767	-3.447	-2.007	-1.897	-2.017	-2.317	-5.967
	TST	-2.087	-2.077	-1.917	-3.437	-2.267	-1.927	-2.147	-2.407	-5.837

*Denotes cases where methods never achieve %C = 90%

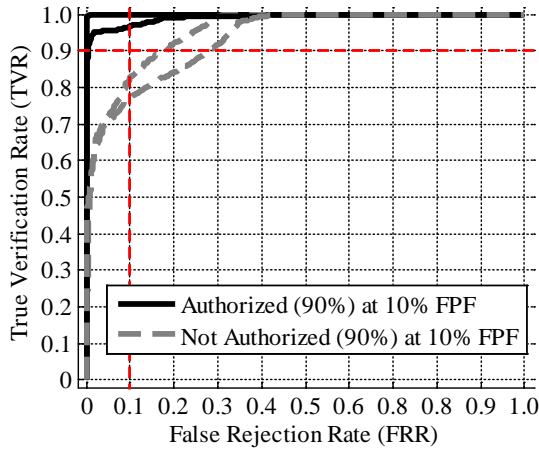
While the *RAP* results in Table IV-7 offer comparable information as seen in Table IV-7, *RAP* enables the ability to examine both $N_F = 10$ performance, which could not be examined using gain, and *RAP* enables a comparison across SNR all operating points. In Table IV-7, higher values indicate higher performance and thus MLF DRA methods are seen to offer the highest performance overall. From a classification standpoint, the loadings methods, especially *SSum*, *UMax*, and *USum* appear to therefore offer higher and more consistent performance. Thus MLF methods offer a clear classification performance improvement over methods previously presented, e.g. [89] [113].

Table IV-7: Relative Accuracy Percentage (RAP) from Baseline $N_{DRA} = 729$ Feature Set. Bold entries with light grey shading denote best case (highest scoring) performance. Reprinted from [135].

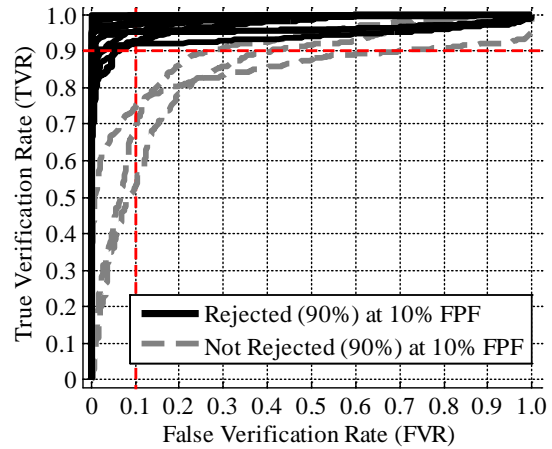
DRA Subset		DRA Feature Selection Method								
		<i>Pre-Classification</i>		<i>Post-Classification</i>						<i>Baseline</i>
		KS	F-Test	GRLVQI	Wilk's	MLF <i>SMax</i>	MLF <i>SSum</i>	MLF <i>UMax</i>	MLF <i>USum</i>	Random
$N_{DRA} = 10$	TNG	65.12	70.82	62.99	71.28	71.12	68.50	71.17	72.71	61.48
	TST	65.52	71.59	63.79	72.29	71.83	68.91	71.84	73.33	61.87
$N_{DRA} = 26$	TNG	78.23	78.14	79.97	77.61	79.38	81.82	79.39	81.85	74.23
	TST	78.99	79.16	80.68	78.69	80.08	82.49	80.04	82.51	74.98
$N_{DRA} = 50$	TNG	87.52	87.25	87.45	85.08	87.59	88.11	87.34	87.42	78.69
	TST	88.05	87.88	88.01	85.95	88.30	88.71	88.17	88.05	79.25
$N_{DRA} = 100$	TNG	92.55	92.44	93.27	90.95	92.93	92.41	92.92	92.01	85.85
	TST	92.86	92.94	93.56	91.51	93.52	92.65	93.56	92.30	86.24
$N_{DRA} = 157$	TNG	94.97	95.54	95.52	92.95	95.47	95.97	95.67	95.59	90.77
	TST	95.39	95.99	95.89	93.37	95.89	96.36	96.16	96.00	91.48
$N_{DRA} = 191$	TNG	96.36	96.69	96.34	94.18	96.41	96.78	96.50	96.76	91.07
	TST	96.70	97.13	96.71	94.54	96.83	97.13	96.87	97.19	91.30
<i>Average RAP</i>		86.02	87.13	86.18	85.70	87.45	87.49	87.47	87.98	80.60
<i>Cumulative RAP</i>		1032.26	1045.57	1034.19	1028.40	1049.37	1049.85	1049.65	1055.71	967.20

4.3.3 DRA Method Verification Performance Assessments

“One vs one” device claimed ID verification performance was considered to further evaluate each DRA classifier model. Figure IV-27a presents authorized device claimed vs. actual ID verification assessment for $UMax$ and $N_F = 50$ at $SNR = 10$ dB, the SNR at which the baseline $N_F = 729$ MDA/ML classifier achieves $\%C = 90\%$ accuracy. The $N_{Auth} = 4$ authorized device ROC curves presented in Figure IV-27a show that 50% of authorized devices are correctly authorized at $TVR > 90\%$ and $FRR < 10\%$ using this model. Figure IV-27b similarly shows the rogue rejection rate for the $UMax$, $N_F = 50$ model at $SNR = 10$ dB. At the threshold of $TVR > 90\%$ and $FVR < 10\%$, 33/36 or 91.7% of *rogue* devices were correctly rejected.



(a) Authorized Based on $TVR > 90\%$ and $FVR < 10\%$ criteria (solid lines), this reflects $TVR = 2/4 = 50\%$ success.



(b) Rogue. Based on $TVR > 90\%$ and $RAR < 10\%$ criteria (solid lines), this reflects $RRR = 33/36 = 91.7\%$ success

Figure IV-27: ZigBee Device ID Verification performance for the $N_{DRA} = 50$ $UMax$ feature subset at $SNR = 10$ dB. Reprinted from [135].

To visually examine the results from the MDA classifiers developed from the DRA algorithms and the DRA assessment methods, a total of 108 ROC curve figure pairs would be needed. Results were therefore generated for all cases and are summarized in

Table IV-8. Here bold entries denote values within 10% of the Best, and bold entries with light grey shading denote best case performance. With the exception of Random selection results, which logically offer the poorest performance for all N_{DRA} subsets, two observations can be made: firstly, that all DRA other selection methods offer comparable verification performance for higher N_{DRA} subsets, e.g. $N_{DRA} = [157, 191]$, and that MLF-based methods generally consistent and generally superior performance for lower dimensional, e.g. $N_F = [10, 26]$, subsets. Consequently the verification performance results concur with the observations seen in the classification results.

Table IV-8: Device ID Verification Performance For $\%C = 90\%$ at $SNR = 10$ dB: True Verification Rate (TVR) for $N_{Auth} = 4$ Authorized Devices and Rogue Rejection Rate (RRR) For $N_{Auth} \times N_{Rog} = 36$ rogue scenarios. Bold entries denote values within 10% of the Best, and bold entries with light grey shading denote best case performance and. Reprinted from [135].

DRA Subset		DRA Method								
		<i>Pre-Classification</i>		<i>Post-Classification</i>						<i>Baseline</i>
		KS	F-Test	GRLVQI	Wilk's	MLF <i>SMax</i>	MLF <i>SSum</i>	MLF <i>UMax</i>	MLF <i>USum</i>	Random
$N_{DRA} = 10$	TVR (%)	0	25	0	25	25	50	25	50	0
	RRR (%)	36.11	52.78	19.44	41.67	38.89	36.11	38.89	50	31.48
$N_{DRA} = 26$	TVR (%)	50	50	50	50	50	50	50	50	25
	RRR (%)	69.44	72.22	80.56	63.89	75	75	77.78	75	51.85
$N_{DRA} = 50$	TVR (%)	50	75	50	75	50	50	50	50	50
	RRR (%)	86.11	91.67	91.67	83.33	91.67	91.67	91.67	88.89	75
$N_{DRA} = 100$	TVR (%)	75	75	100	75	75	75	75	75	66.67
	RRR (%)	94.44	94.44	94.44	94.44	94.44	94.44	94.44	94.44	86.11
$N_{DRA} = 157$	TVR (%)	100	100	100	100	100	100	100	100	75
	RRR (%)	94.44	94.44	94.44	94.44	94.44	94.44	94.44	94.44	91.67
$N_{DRA} = 191$	TVR (%)	100	100	100	100	100	100	100	100	75
	RRR (%)	97.22	97.22	94.44	94.44	97.22	97.22	97.22	97.22	91.67

V. Extensions to the LVQ-Family of Algorithms

The ant, viewed as a behaving system, is quite simple. The apparent complexity of its behavior over time is largely a reflection of the complexity of the environment in which it finds itself.

—HERBERT A. SIMON, 1916-2001

While various studies have extended Learning Vector Quantization (LVQ) algorithms by considering non-Euclidean distance measures, the extensions are not always correctly formulated and the reason(s) for considering alternate measures is not always clear. Below, the Generalized Relevance Learning Vector Quantization Improved (GRLVQI) process is fundamentally extended via a process to select and incorporate alternative distance measures. As discussed in Chapter III, differences in LVQ algorithms generally revolve around cost functions and hence changing distance measures involves deriving new update equations.

5.1 Introduction

Herein, overall LVQ algorithm considerations include the following:

- 1) a minor general improvement to LVQ algorithms is made by using a scaled gradient descent which enables direct comparison of learning rates between problems;
- 2) approaches for selecting the number of Prototype Vectors (PVs) are considered;
- 3) a derivative skeleton framework is created to generalize the process for incorporating alternate distance measures into LVQ, Relevance Learning

Vector Quantization (RLVQ), Generalized Learning Vector Quantization (GLVQ), Generalized Relevance Learning Vector Quantization (GRLVQ) and GRLVQI algorithms;

- 4) a methodology is formalized for proper selection and incorporation of distance measures and learning rates;
- 5) a new cost function is presented for GLVQ, GRLVQ, and GRLVQI algorithms to permit a wide variety of distance measures to be considered;
- 6) a design of experiments (DOE) methodology with Analysis of Variance (ANOVA)-based response surface methods and optimization of algorithm parameter settings through sequential quadratic programming (SQP) are employed to find optimal operating points. The primary benefit of these improvements is that finding appropriate algorithm parameter settings is optimized and a systematic process for deciding which distance measure to use in LVQ algorithms is developed and considered.

The resultant improved GRLVQI algorithm is termed GRLVQI-Distance (GRLVQI-D) to indicate the algorithm is generic and can be adopted to use any differentiable distance measure. Additionally, similar extensions to the GLVQ and GRLVQ algorithms are made with these extended algorithms termed GLVQ-D and GRLVQ-D, respectively.

This chapter is organized as follows. Firstly, algorithmic development aspect relative to LVQ through GRLVQI are presented in Section 5.2. The GRLVQI-D algorithm is presented in Section 5.2.2.4 and a procedure is developed and applied in

Section 5.3 for selecting distance measures for GRLVQI-D. GRLVQI-D is extended to RF-DNA Fingerprinting in Section 5.4.

5.2 GRLVQI-D Algorithm Development

High levels of dimensionality are known to adversely affect Euclidean distance based classifiers [470, 471], which is directly relevant to RF-DNA applications of LVQ algorithms since RF-DNA fingerprint features generally have a large number of features and exemplars. Therefore, incorporating a non-Euclidean distance metric in GRLVQI could be advantageous. However, to incorporate a non-Euclidean distance measure the underlying cost-function must be changed in a given LVQ algorithm.

5.2.1 Prior Implementations of non-Euclidean Distances in LVQ

In LVQ algorithms, a gradient descent is used with the step size a function of the cost function. A gradient descent implicitly requires evaluating the gradient of the associated cost function; therefore, a new PV update expression must be computed for any change in the distance equation or cost function. GRLVQ and GRLVQI were developing using squared Euclidean distance for selecting prototype vectors [245]. Other LVQ variations have seen improvement through difference distance metrics, e.g. the innovations of Schneider et al. [298] to where two new metrics similar in form to Mahalanobis distances were incorporated into GRLVQ.

Common issues in LVQ distance measure extensions is neglecting to compute a new PV gradient descent update equation when considering alternative distance equations

and incorrect formulations, c.f. [472–475]. These common pitfalls found in the LVQ literature. PV update equations can be generalized, per Ji et al. [476, 477], as

$$\mathbf{w}(t + 1) = \mathbf{w}(t) + c\mathbf{x}, \quad (5.1)$$

where c is a scalar and \mathbf{x} is the PV update. However, such formulations imply that c is merely a scalar step size when in fact it is composed of both the learning rate and a gradient descent specified quantity. This is an important distinction since any given c is specific to the cost function, learning rate, and the distance equation employed.

Biehl et al. [290] created distance measure variants for GRLVQ; however, the process presented in Biehl et al. is not easily generalizable to other distances and the equations are presented with non-intuitive formulations. Strickert et al. [291] formulated a GRLVQ variant using a correlation based measure and provided justification for using both distance metrics and measures; however, the formulation skipped over multiple steps to make it generalizable to other problems. When a different distance measure is used direction of the PV update must be considered relative to the direction of the distance measure [291]. The solution adopted herein and suggested by Strickert et al. [291] is to merely flip the signs on the PV update equations [291].

However, all of these approaches created specific formulations and were not readily generalizable. Since, the process and equations presented for these applications is not always intuitive or correctly followed, creating a general framework to facilitate formulating PV update equations is beneficial. To create such a framework, the process used to formulate PV update equations must be understood and components identified

that need to be changed whenever a new distance equation is to be used. Therefore, to avoid any confusion, the entire PV update equation is reported herein.

5.2.2 Developing a Differentiation Skeleton for LVQ Improvements

The following general improvements are made to LVQ algorithm. First, Section 5.2.2.1 presents a scaled gradient descent method for any LVQ algorithm to enable direct comparison of learning rates. Then Section 5.2.2.2 discusses gradient descent considerations when making changes to LVQ algorithms, supporting derivations are provided in Appendices E and F. Cost function extensions to GLVQ, GRLVQ, and GRLVQI are discussed in 5.2.2.3 and Appendix G. Finally, relevance derivatives for GRLVQ and GRLVQI algorithms are discussed in 5.2.2.3 and Appendix H. A differentiation skeleton for incorporating any differentiable distance measure in LVQ, RLVQ, GLVQ, GRLVQ, and GRLVQI is then presented in 5.2.2.4.

5.2.2.1 Scaled Gradient Descent

Widrow-Hoff (W-H) learning is a least mean squares formulation for the gradient descent [243; 250, pp. 55-57; 478]. W-H considers a squared Euclidean distance metric (e) for general gradient descent updating of LVQ [250, pp. 55-57; 478]. The gradient of function f is given by

$$\nabla f_K = \left(\frac{\partial f_K}{\partial X_1}, \frac{\partial f_K}{\partial X_2}, \dots, \frac{\partial f_K}{\partial X_p} \right), \quad (5.2)$$

where K is the step number and p is the number of variables [243]. From (5.2), a gradient search for a maximum can be computed via

$$\tilde{X}_{i+1} = \tilde{X}_i + \frac{\delta \nabla f_i}{\|\nabla f_i\|}, \quad (5.3)$$

where δ is the learning rate or step size [243]. Given $\delta/\|\nabla f_i\|$ is a scalar, the scaled learning rate can be incorporated in other gradient descents. Considering the gradient descent algorithm in (3.20), it can be rewritten as

$$\mathbf{w}_e(t+1) = \mathbf{w}_e(t) + \epsilon^*(t) \nabla e, \quad (5.4)$$

where,

$$\epsilon^*(t) = \frac{\epsilon(t)}{\|\nabla e\|}. \quad (5.5)$$

The underlying advantage of incorporating (5.4) and (5.5) in LVQ, RLVQ, GLVQ, GRLVQ and GRLVQI is that it enables a direct comparison of learning rates across LVQ methods and datasets without significantly changing the algorithms.

5.2.2.2 Gradient Descent and Derivatives in LVQ Algorithms

To incorporate a non-Euclidean distance measure in LVQ, we must consider the gradient computation, as seen in (3.20) and discussed in Section 3.3.1, of the cost function $C(\mathbf{w}_n(t))$. For LVQ, the cost function is the distance measure itself. Therefore, creating a non-Euclidean distance LVQ algorithm requires 1) selecting a distance measure to replace (3.21), and 2) updating the cost function by computing the first derivative of the new measure to replace $x_i - \mathbf{w}_e(t)$ in (3.24) and (3.25). The appropriate in-class PV signs would then be computed per the derivative and then considered with respect to what the new measure represents.

(a) **Gradient Descent in RLVQ Relevance Computation**

Per the discussion in both Section 3.3.1.4 and [266], the RLVQ expression in (3.31) is also computed via a gradient descent. Thus, when changing a distance metric in RLVQ it is necessary to change the cost function. When considering the RLVQ gradient descent in (3.29), the cost function for RLVQ is the distance in (3.30). The product rule for derivatives is,

$$d(uv) = u dv + v du \quad (5.6)$$

where u and v are two different variables [279]. For the RLVQ cost function, one logical choice would be $u = \psi$ and $v = (x_i - w_n)^2$, which is considered for $\partial d / \partial \psi$, the derivation of the distance d with the respect to ψ . This results in the following derivation:

$$\begin{aligned} \frac{\partial d}{\partial \psi} &= \psi \cdot 0 + 1 \cdot (x(t) - w(t))^2 \\ &= (x(t) - w(t))^2 \end{aligned} \quad (5.7)$$

with the final expressing being the expression in (3.31) with the sign being associated with convention where smaller values indicate higher significance and larger values indicate lower significance [266].

(b) **Gradient Descent in GLVQ, GRLVQ, and GRLVQI**

Although the gradient descent derivations for LVQ and RLVQ appear trivial, as discussed in Section 5.2.2.2(a), the derivations are non-trivial when the gradient descents are computed for GLVQ, GRLVQ and GRLVQI. To fully understand the process,

derivations for the PV update gradient descent operations and relevance gradient descent are discussed in Appendices E through H.

5.2.2.3 Cost Function Extensions for the GLVQ Family of Algorithms, the GLVQ-D, GRLVQ-D, and GRLVQI-D Algorithms

The nominal relative distance difference equation for GLVQ, GRLVQ, and GRLVQI presents issues when non-Squared Euclidean distance measures are used. For this equation to yield the expected values between -1 and $+1$, it assumes that the distance measure yields a positive value. When changing the distance measure to a non-squared Euclidean distance one is not ensured of the distance being positive. Hence selecting an appropriate relative distance difference equation is necessary. Two obvious approaches were considered: an absolute value measure, where the absolute value of each distance is taken, and a squared measure, where each distance is squared. The absolute value approach, which would consider

$$\mu(x^m) = \frac{(|d^J| - |d^L|)}{(|d^J| + |d^L|)}, \quad (5.8)$$

has notable issues and was not developed further because this would require an overly complex gradient descent method due to there being three conditions of absolute value derivatives: positive, negative, and 0 when the function itself is continuous but not differentiable at 0 [479]. Therefore, only an improved squared relative distance difference function will be developed and considered.

In order for the new relative distance difference equation to compute the same scores for the nominal squared-Euclidean distance measure, the following improved equation was developed,

$$\mu(x^m) = \frac{(d^J)^2 - (d^L)^2}{(d^J)^2 + (d^L)^2}, \quad (5.9)$$

where each distance is ensured to be positive. However, by changing $\mu(x^m)$ a new GLVQ gradient descent must necessarily be computed, per Section 5.2.2.2(b). The derivation for the new GLVQ gradient descent is presented in Appendix G, with the resultant PV update becoming

$$\begin{aligned} \mathbf{w}^J(t+1) &= \mathbf{w}^J(t) + \frac{8\epsilon(t)(\partial f / \partial \mu(x^m))d^L}{(d^J + d^L)^2} (\mathbf{x}^m - \mathbf{w}^J)^3 \\ \mathbf{w}^K(t+1) &= \mathbf{w}^L(t) - \frac{8\epsilon(t)(\partial f / \partial \mu(x^m))d^J}{(d^J + d^L)^2} (\mathbf{x}^m - \mathbf{w}^L)^3. \end{aligned} \quad (5.10)$$

which differs from the PV updates in (3.35) only by the scalar multiplier and the squared terms in the relative distance difference equations.

When considering GRLVQ or GRLVQI, one must also update the relevance gradient descent if the relative distance difference equation has been changed. Appendix H presents this process for (5.9) and yields a new relevance update,

$$\begin{aligned} \psi_q(t+1) &= \psi_q(t) \\ &- \epsilon(t)f'|_{\mu(x^m)} \left(-\frac{2(d^J - d^L)(x_{iq}(t) - w_{nq}(t))^2}{(d^J + d^L)^2} \right) \end{aligned} \quad (5.11)$$

which is equivalent to the GRLVQ relevance update in (3.37) prior to being multiplied and written out. Following the considerations of Section 3.3.1.6 and Appendices E through H, the underlying GRLVQ gradient descent PV gradient descent is thus,

$$\begin{aligned}
\mathbf{w}^J(t+1) &= \mathbf{w}^J(t) + \frac{8\epsilon(t)(\partial f/\partial \mu(x^m))d^L}{(d^J + d^L)^2} \boldsymbol{\Psi} \cdot (\mathbf{x}^m - \mathbf{w}^J)^3 \\
\mathbf{w}^K(t+1) &= \mathbf{w}^L(t) - \frac{8\epsilon(t)(\partial f/\partial \mu(x^m))d^J}{(d^J + d^L)^2} \boldsymbol{\Psi} \cdot (\mathbf{x}^m - \mathbf{w}^L)^3.
\end{aligned} \tag{5.12}$$

5.2.2.4 A Differentiation Skeleton for LVQ Distance Metrics

Examining the derivation process that yields the PVs updates for LVQ, RLVQ, GLVQ, or GRLVQ, one can notice a few patterns. Firstly, while the gradient descent cost function in LVQ and GLVQ differs dramatically, one will compute the same first derivative for a given distance metric for both algorithms since the distance metric is the cost function in LVQ, per (3.20)–(3.25). In GLVQ, the distance metric first derivatives are the same as in LVQ except for denotation for the appropriate in-class and out-of-class distances, (E.17)–(E.20), however additional derivatives must be computed for the cost function, (E.1)–(E.5), and the relative distance difference equation, (E.7)–(E.16). These must then be assembled; however, these are noticeably identical when changing distance measures except for (possibly) sign and the appropriate in/out of class subscript. Additionally, as long as the same logistic sigmoid cost function is employed per (E.1)–(E.5) then one does not need to recompute its derivative, $f'(\mu(x^m))$. Similarly, the derivatives in RLVQ and GRLVQ are closely related to the derivative computed for their respective cost function.

As long as the underlying gradient descent process in (3.20) is not changed, the derivative approach will be consistent. It is intuitively obvious to the casual observer that as long as both the difference equation in (3.34) is used, then general quotient rule

process in (E.6) will be consistent and therefore changing the distance metric in a GLVQ type of gradient descent process merely involves computing the following derivatives du/dw^J , dv/dw^J , du/dw^L and dv/dw^L and then only computing the resultant equation via the quotient rule.

Following the above knowledge, Figure V-1 presents decomposition of GLVQ, GRLVQ and GRLVQI gradient descents and from where each respective part is computed. Using this knowledge, one can determine which component of the gradient descent needs to be updated based upon which change in the algorithm. For example, if only the distance measure is changed, then only the component in red needs to be changed; care must be taken with the scalar multiplier, since this is a function of both the distance measure and relative distance difference, and it could further also be a function of the cost function, depending on what is changed.

Observable in Figure V-1 is that this visualization is generalizable to LVQ as well as GLVQ algorithms. For instance, in LVQ and RLVQ, the cost function of the gradient descent is the distance measure itself and thus the distance measure and relative distance difference measure related components of Figure V-1 are not considered and one only computes the derivative of cost function. One can further similarly observe relevance updates as seen in Figure V-2. Extending from these observations, an algorithmic skeleton for making various changes to LVQ, RLVQ, GLVQ, GRLVQ, and GRLVQI is presented in Figure V-3.



$$w^{J,K}(t+1) = w^{J,K}(t) \pm \frac{\overset{\text{distance measure related}}{4} \overset{\text{cost function related}}{\varepsilon(t)} \overset{\text{relative distance difference metric related}}{f' \big|_{\mu(x^M)}} \overset{\text{relative distance difference metric related}}{d^{K,J}}}{\overset{\text{relative distance difference metric related}}{(d^J + d^K)^2}} \overset{\text{distance measure related}}{(x^m - w^{J,K})}$$


Figure V-1: Components of GLVQ, GRLVQ and GRLVQI Gradient Descents.

$$\psi(t+1) = \psi(t) - \xi(t) \overset{\text{cost function related}}{f' \big|_{\mu(x^M)}} \left(\frac{\overset{\text{relative distance difference metric related}}{d^K}}{\overset{\text{relative distance difference metric related}}{(d^J + d^K)^2}} \overset{\text{distance measure related}}{(x^m - w^J)^2} - \frac{\overset{\text{relative distance difference metric related}}{d^L}}{\overset{\text{relative distance difference metric related}}{(d^J + d^K)^2}} \overset{\text{distance measure related}}{(x^m - w^K)^2} \right)$$


=distance measure related

=cost function related

=relative distance difference metric related

Figure V-2: Components of GLVQ, GRLVQ and GRLVQI Gradient Descents.

Algorithm 3 LVQ Derivative Framework

```

Select new distance metric  $d(x, w)$ 
  if  $\partial d(x, w)/\partial w$  exists do
    Compute  $\nabla C(w(t)) = \partial d(x, w)/\partial w$ 
    Insert  $\nabla C(w(t))$  into LVQ algorithm per  $w(t+1) = w(t) - \epsilon(t)\nabla C(w(t))$ 
    Use new  $d(x, w)$  in  $\arg \min_i (\sum d(x_i, w_i))$ 
  end
  if RLVQ expression desired
    Extend  $d(x, w)$  function to include relevance
    Compute  $\partial d(x, w)/\partial \psi$ 
    Extend LVQ function to include logic for relevance computation
  end
  if GLVQ expression desired
    Select cost function,  $f(\mu(x^m))$ , and distance measure  $\mu(x^m)$ 
    Compute derivative for cost function  $f(\mu(x^m))$  via
      
$$\frac{\partial f(\mu(x^m))}{\partial w} = \frac{\partial f(\mu(x^m))}{\partial \mu(x^m)} \frac{\partial \mu(x^m)}{\partial w}$$

    Compute derivative for sigmoid:
      
$$\frac{\partial f(\mu(x^m))}{\partial \mu(x^m)} = f'(\mu(x^m))\mu'(x^m)$$

    Consider sigmoid distance metric and compute for
       $\partial \mu(x^m)/\partial w^J$  &  $\partial \mu(x^m)/\partial w^L$ 
    if  $\mu(x^m) = \frac{(d^J - d^L)}{(d^J + d^L)}$ 
      Compute:
      
$$\frac{\partial \mu(x^m)}{\partial w^J} = \frac{\partial \mu(x^m)/\partial w^J (2d^J)}{(d^J + d^L)^2} \text{ and } \frac{\partial \mu(x^m)}{\partial w^L} = \frac{\partial \mu(x^m)/\partial w^L (-2d^L)}{(d^J + d^L)^2}$$

    else
      Compute new derivative expression for distance measure
    end
    Assemble equations
  end
  if GRLVQ or GRLVQI expression desired
    Follow procedure for GRLVQ
    Compute  $\partial d(x, w)/\partial \psi$ 
    Assemble equations
  end
end

```

Figure V-3: Pseudocode Process and Derivative Skeleton for Changing Distance Metrics in LVQ, RLVQ, GLVQ, and GRLVQ.

5.3 Selecting Distance Measures for GRLVQI-D

With the GRLVQI-D algorithm formalized, one must now determine which distance measure should be incorporated. However, the process presented in Section 5.2.2 being formalized, it is still non-trivial considering the various derivatives and computations. It is additionally, non-intuitive on which distance measure to select. Appendix I reviews various distance measures as described by Cha [283] in his review of distance measures.

A general distance measure selection process for LVQ algorithms is therefore presented due to 1) the long list of possible distance measures, 2) the involved derivation process required to implement a new distance measures into GRLVQ or GRLVQI, 3) the large amount of data and computation time needed for RF-DNA applications, and 4) no extant guidance on which distance measures should be considered. The proposed distance measure selection process innovates via the following, 1) distance measures are first compared via correlation on two random vectors, 2) uncorrelated distance measures then are then selected via statistical clustering, then 3) the gradient, first derivatives, of these measure are computed and LVQ performance is examined on an academic problem dataset, and finally, 4) measures that offer good performance in LVQ are then examined in RLVQ, GLVQ, and GRLVQ. Underperforming distance measures are not considered in subsequent algorithms, e.g. a measure that performs poorly in LVQ is not considered in RLVQ, due to the general belief that if one cannot solve a simple problem then one will have difficulties solving more complex problems. Figure V-4 presents the general

methodology for selecting distance measures and developing distance measure variants of GRLVQI.

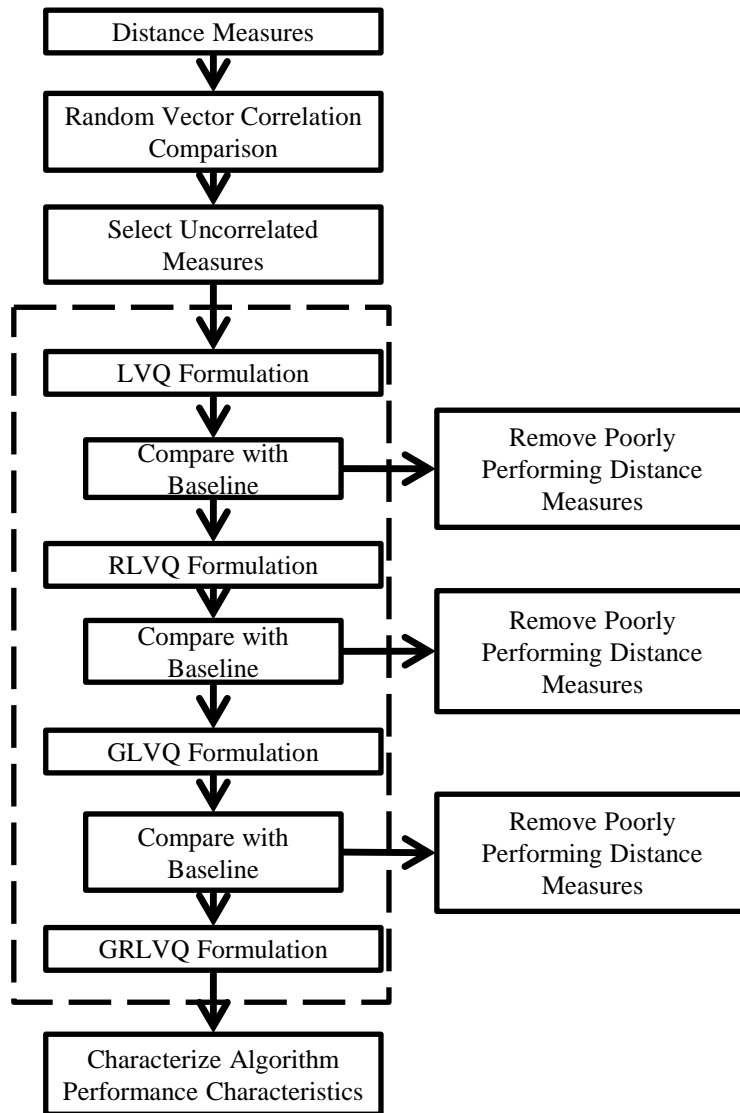


Figure V-4: Iterative Process for Selecting Distance Metrics for GRLVQI.

5.3.1 Selecting Distance Measures for Consideration

Cha [283] identified 62 different distance measures and metrics, which can be grouped into 9 related groups as described in Appendix I: Minkowski, L_1 , Intersection, Inner Product, Fidelity, Squared L_2 , Shannon's entropy, Combinations, and Vicissitude. However, many of these distance metrics are highly related, correlated, or contain non-differentiable factors. Therefore, only a few were evaluated for GRLVQI and measures employing maximization or minimization were not considered due to the dubious derivations [480]. Considering the excluding factors, 22 measures remained for consideration: Euclidean, City Block, Squared Euclidean, Sorensen, Canberra, Inner Product, Harmonic Mean, Cosine, Pseudo-Cosine, Kumar-Hasselbrook, Jaccard, Dice, Pearson χ^2 , Neyman χ^2 , Squared χ^2 , Divergence, Additive Symmetric, Kumar-Johnson, Covariance, Correlation, Mahalanobis, and Squared Mahalanobis.

5.3.2 Comparing Potential Distance Metrics via Correlation

To understand how the remaining 22 distance measures were related, a correlation study was posed where distance measures are grouped based upon correlation of results and only dissimilar distance measures are selected for further analysis for incorporation into LVQ. To quantify the correlation between distance measures, two uncorrelated random normal vectors of length 1,000 were permuted. These vectors were uncorrelated with a Pearson correlation coefficient of 0.024. These uncorrelated vectors were then inserted for P and Q in the appropriate equations seen in Appendix I, and then 1,000 paired distances between P and Q were then computed.

Figure V-5 presents a correlation matrix between the paired distance measures results. A few observations can be made from Figure V-5, firstly, many distance metrics are highly correlated only within Cha's [283] 'families' or groups; secondly, there no measure appears highly correlated with all other measures; and thirdly, both positive and negative correlations are seen. Positive and negative correlations should logically be considered with respect to the nominal squared Euclidean measure; measures that are negatively correlated with the squared Euclidean measure logically have larger values for more similar exemplars and smaller values for more different exemplars, consistent with [481], when employing measures negatively correlated to Squared Euclidean distance one desires to maximize the distance rather than minimize.

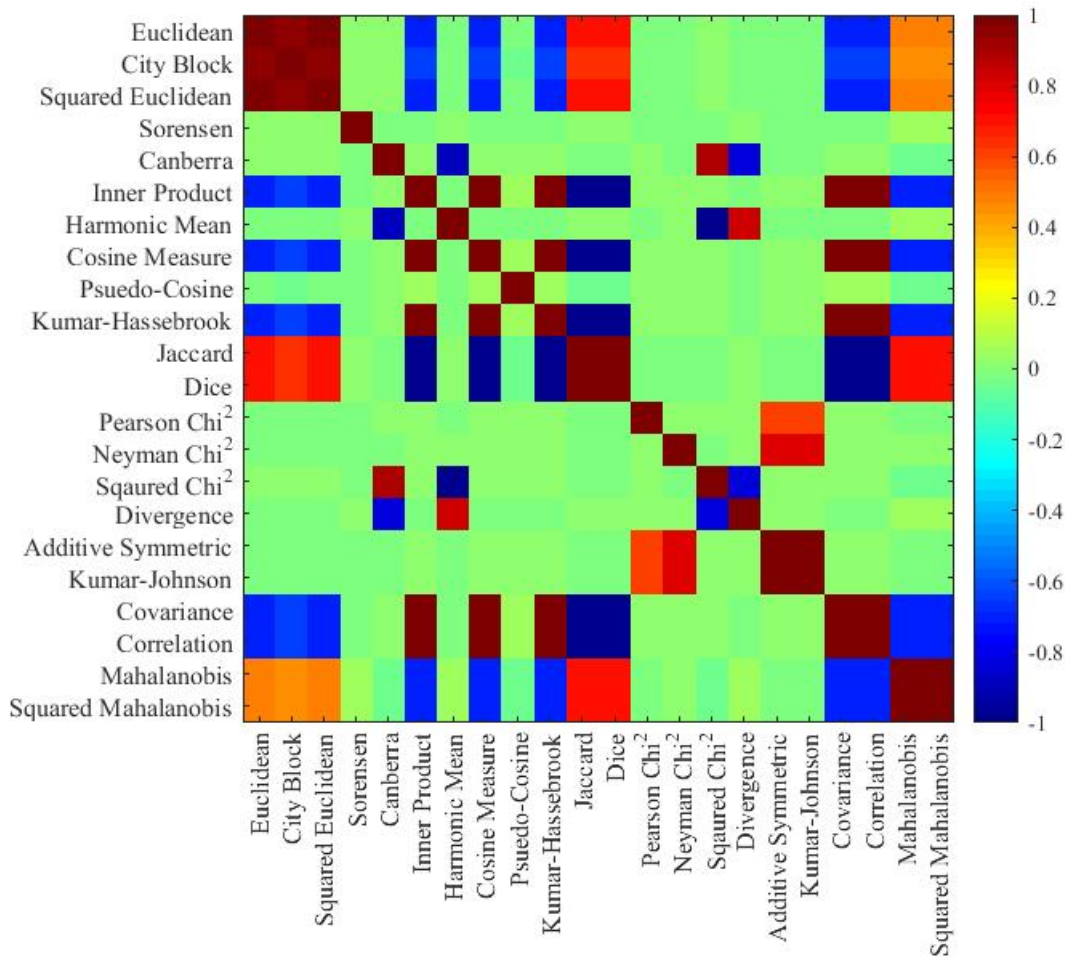


Figure V-5: Correlation Comparison of Distance Metrics on Random Normal Data.

To select distance measures for inclusion into GRLVQI hierarchical clustering, consistent with [482], was used to find groups of distance measures. Hierarchical clustering considers a distance matrix between variables and then applies a linkage approach to determine how variables are connected [448]. For a distance matrix, the correlation matrix from Figure V-5 was used since this is the relative distance of interest.

A dendrogram, a diagram employed in cluster analysis to show partitions and

closeness of variables [236], is presented in Figure V-6. Figure V-6 is viewed and interpreted as follows: the y-axis indicates closeness of variables, and ranges from 0 (similar) to a maximum of 4 (distant) [449]. At the maximum value, all variables are linked together, heading towards zero (where only similar variables are linked) groups are determined through an appropriate linkage method [449, 450]. The complete linkage method, which finds most distant pairs and groups less distance pair together [236, 451], was used to evaluate closeness.

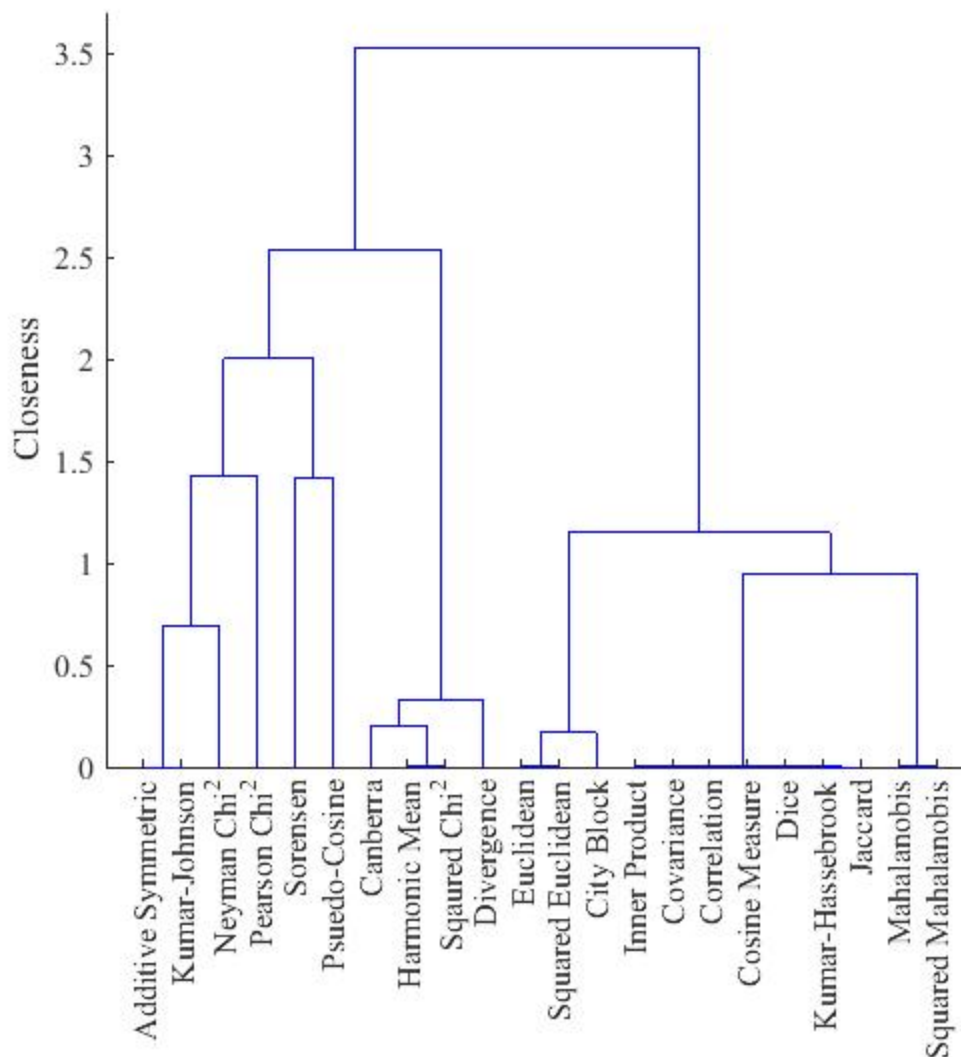


Figure V-6: Dendrogram with Complete Linkage and Correlation Matrix, from Figure V-5, as Distance Matrix.

The number of clusters, and hence number of distance measures to consider, was determined by setting a subjective closeness threshold by considering how far apart the groupings in Figure V-6 appear. A threshold of 0.5 was used, resulting in 9 clusters to consider. A “Chinese Menu” approach, consistent with [486–491], was then used to

select distance measures wherein one method from each group was selected. To facilitate derivations and inclusion into LVQ algorithms, the simplest distance equation in each group was selected. This resulted in the following nine distance measures being further considered: Additive Symmetry, Neyman χ^2 , Pearson χ^2 , Sorensen, Pseudo-Cosine, Canberra, Squared Euclidean, Cosine, and Squared Mahalanobis.

5.3.3 Determining Suitable Distance Measures and LVQ Algorithm Settings

To understand how LVQ distance measure extensions behave for various operating points, a small academic dataset was considered and learning and relevance rates were considered for each LVQ distance measure variant. As underperforming algorithms were found, they were not considered further, e.g. poorly performing LVQ distance measure variances were not further considered in RLVQ. Fisher Iris [235], a small academic dataset, was considered with $N_F=4$, $N_{obs} = 150$, with data equally divided among $N_C = 3$ classes. Training and testing sets were segregated by taking the first 45 observations from each class for training with the remaining 5 observations per class considered as testing. To remove randomization issues, 100 iterations were considered with the classification accuracy averaged.

Because the dynamic range and values computed by the different distance metrics will differ, before considering RF-DNA data in GRLVQI first the relationship between learning rate and number of PVs was explored in LVQ with the Fisher Iris academic dataset. This provides an understanding of how each measure behaves and how each measure behaves compared to the nominal squared-Euclidean distance metric. This

approach is considered iteratively, as described in Figure V-3, with each measure first examined in LVQ, then RLVQ, GLVQ, and finally, GRLVQ. As measures are found to offer little or no performance benefits, they are removed from consideration in further iterations (e.g. if a given measure performs poorly in LVQ, it is not examined in RLVQ, GLVQ, or GRLVQ) since, logically, if a measure offers poor performance and relatively little understanding of its behavior in a simple algorithm it will be difficult for it to offer good performance in a complex algorithm.

In each algorithm the normalized learning and relevance rates were considered for 8 different levels as presented in Table V-1. These settings provide various conditions around the nominal LVQ settings, as described in Section 3.2.1.8. For RLVQ and GRLVQ each combination of learning and relevance rate was explored.

Table V-1: Learning and Relevance Rates for LVQ Algorithm Experiment.

Level	Learning Rate	Relevance Rate
1	0.0001	0.0001
2	0.001	0.001
3	0.01	0.01
4	0.1	0.1
5	1.0	1.0
6	10.0	10.0
7	100.0	100.0
8	1000.0	1000.0

5.3.3.1 Determining Suitable Distance Measures and LVQ Algorithm Settings

Figure V-7 presents results after formulating the LVQ cost functions, provided in Appendix J, and computing performance results for each LVQ variation. As seen in Figure V-7, only 5 LVQ distance measure variants achieve training or testing classification above 40%. Squared Euclidean (the baseline), Cosine LVQ, and Canberra LVQ consistently perform above 60% accuracy for learning rates above 0.1 and thus these methods will be further explored for other LVQ variations. While Neyman χ^2 and Sorensen LVQ variants achieve between 40 and 60% classification accuracy, they perform much worse than Squared Euclidean, Cosine and Canberra and thus Neyman χ^2 and Sorensen LVQ variants are not considered further.

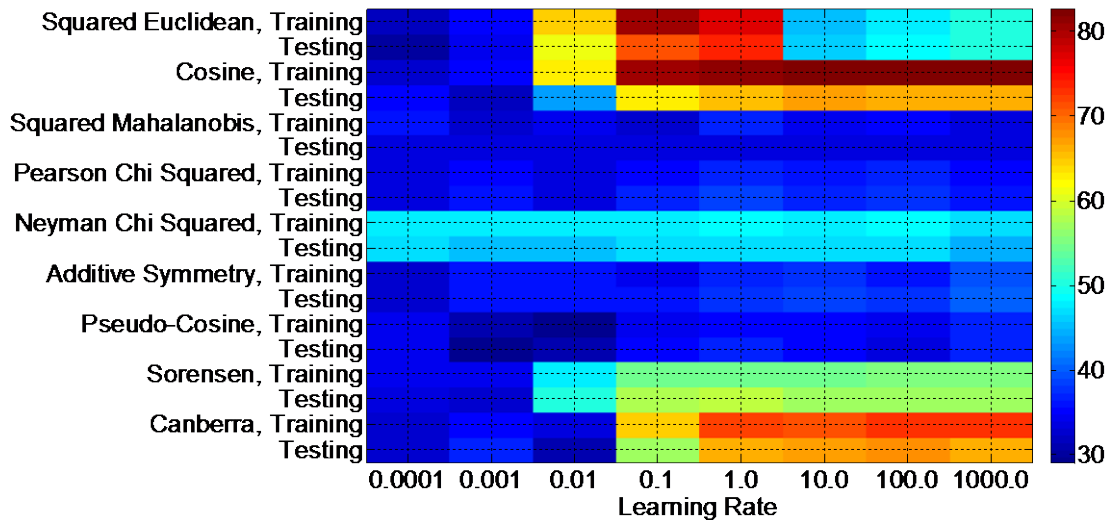


Figure V-7: Distance Measure Performance versus Learning Rate for LVQ

5.3.3.2 Distance Measure Extensions to GLVQ

As discussed in Section 5.2.2.4, changing the distance measure in the cost function for GLVQ involves merely changing the distance measure component of the cost function derivative. This was considered for Squared Euclidean (baseline), Cosine, and Canberra measures. Figure V-8 presents classification results, best performance is seen for learning rates of 0.01 and 0.1 for Squared Euclidean, above 1.0 for Cosine, and at 0.1 for Canberra. Thus one could interpret this as indicating that Cosine GLVQ needs a learning rate 10-100 times that of Squared Euclidean to achieve reasonable performance.

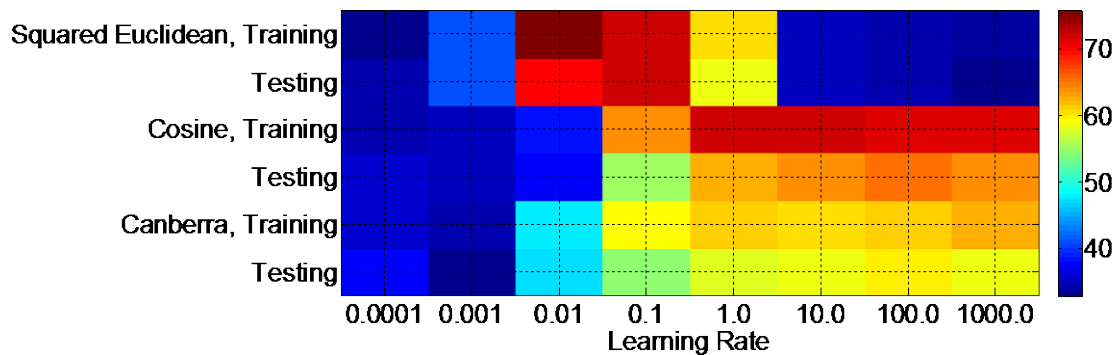


Figure V-8: Distance Measure Performance versus Learning Rate for GLVQ

5.3.4 Relevance Learning with Alternative LVQ Distance Measures

Care must be taken when incorporating relevance learning in distance measures since the relevance weighting must be relative to each feature. In RLVQ, the Euclidean distance measure of (3.21) is formulated so that the relevance multiplier is easily contained inside the summation. However, it is not always obvious where to

incorporating the relevance multiplier on different distance measure, such as both the Canberra and Cosine measures.

The Canberra measure consists of a summation of two ratios; to ensure the relevance values are associated with features and not PVs, the relevance values must therefore be a Hadamard product, e.g. [492], to ensure appropriate weighting on each feature. Although Sorensen was not considered beyond LVQ, its formulation as a ratio of sums would increase difficulties in incorporating relevance learning. To implement relevance learning, the relevance must be added so that it multiplies to each feature, for Canberra the following relevance distance measure appropriately accomplishes this,

$$d_{can,\psi} = \sum_{i=1}^{N_F} \psi_i \frac{x_i - w_i}{x_i + w_i}. \quad (5.13)$$

When considering the Cosine distance measure, one sees a summation of a ratio with the numerator being a product and the denominator a product of two summations. To avoid an overly complicated derivative the relevance multiplier was added to only the numerator, with the Cosine relevance equation appearing as,

$$d_{cos,\psi} = \sum_{i=1}^{N_F} \frac{\psi_i x_i w_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n w_i^2}}. \quad (5.14)$$

After incorporating relevance learning into RLVQ, using the formulations described in Appendix J and Figure V-3, each algorithm was considered for all relevance rates in Table V-1 and learning rates associated high accuracy (%C > 60%) from Section 5.3.3.2. Classification results are presented in Figure V-9 through Figure V-11 which

shows the relationship between learning rates and relevance rates for Squared Euclidean, Canberra, and Cosine RLVQ algorithms.

Figure V-9 presents the relationship between classification accuracy, learning rates and relevance rates for Squared Euclidean RLVQ on the Fisher Iris dataset. Evident in Figure V-9 is that the best performance is seen when the relevance rate is equal to or less than the learning rate, consistent with [291]. Similarly, Figure V-10 presents Canberra-RLVQ results where the best performance is seen when relevance rate is less than the learning rate and particularly when the relevance rate is equal to 0.01 or less. Finally, Figure V-11 presents classification results for Cosine-RLVQ wherein one sees that the best performance is only achieved when the relevance rate is less than the learning rate and valued 0.0001 or 0.001.

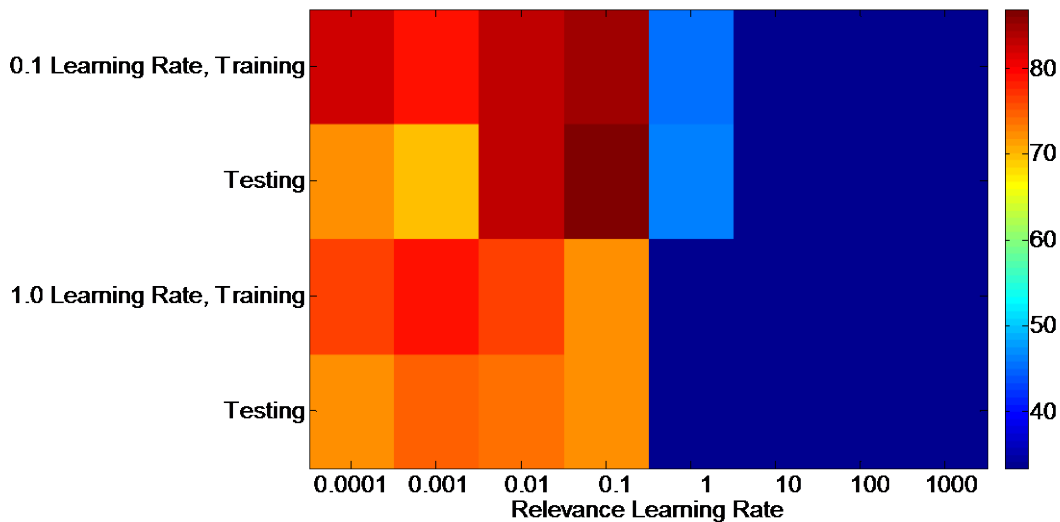


Figure V-9: Learning Rate vs Relevance Learning Rate for Squared Euclidean RLVQ

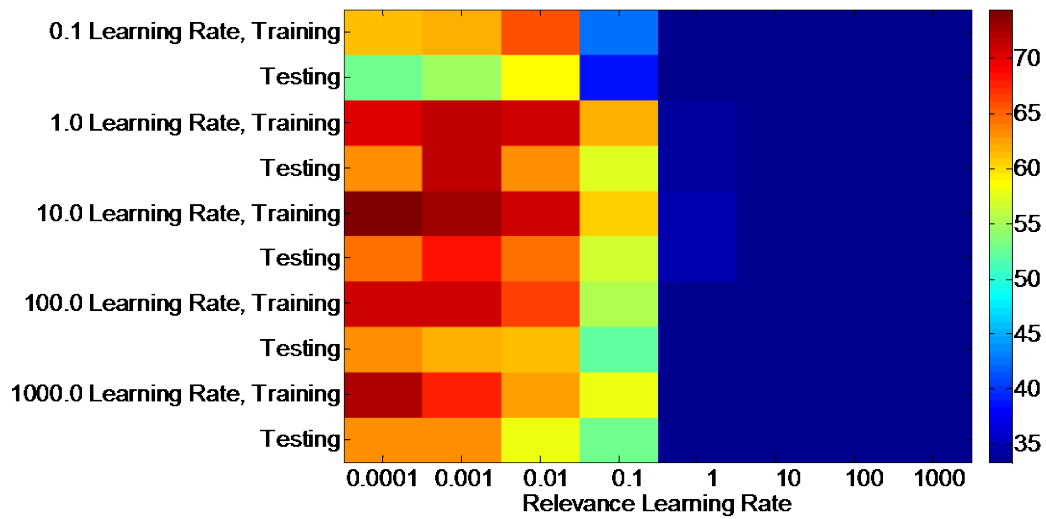


Figure V-10: Learning Rate vs Relevance Learning Rate for Canberra RLVQ

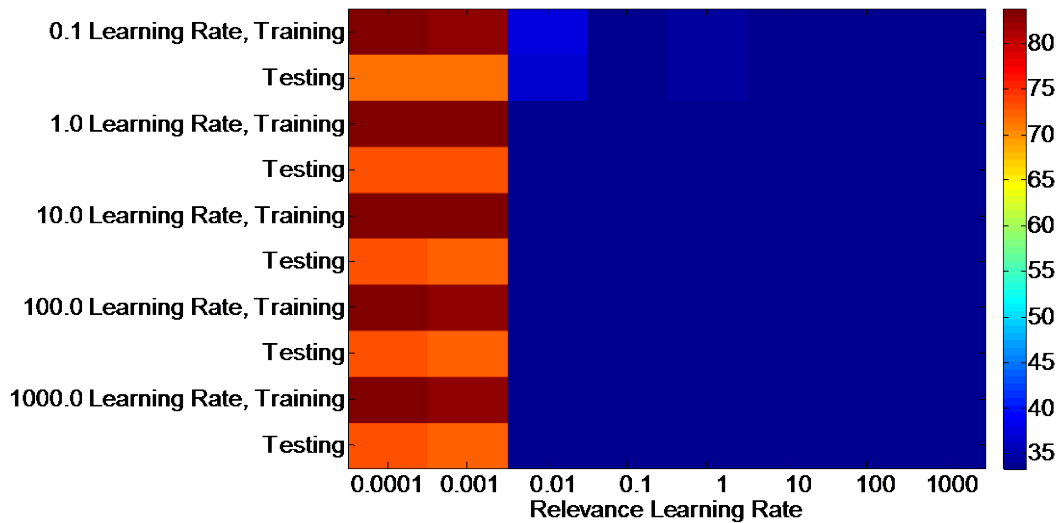


Figure V-11: Learning Rate vs Relevance Learning Rate for Cosine RLVQ

5.3.5 Distance Measure Extensions to GRLVQ and GRLVQI

To extend Canberra-GLVQ and Cosine-GLVQ to include relevance, the considerations of the process in Figure V-3 and Figure V-4 were applied with the GLVQ

sigmoidal cost function and the revised relative distance difference metric of Section 5.2.2.3.

5.3.5.1 Relevance Learning and GRLVQ Extensions

When extending the distance measure formulations to GRLVQ, the considerations described in Figure V-3 were followed wherein the distance measure versions of GLVQ were extended with relevance logic. Figure V-12 presents the relationship between classification accuracy, learning rates and relevance rates for Squared Euclidean GRLVQ on the Fisher Iris dataset. Consistent with Squared Euclidean GRLVQ in Section 5.3.4, evident in Figure V-12 is that the best performance is seen when both the learning rate is less than 1.0 and the relevance rate is less than the learning rate. Similarly, Figure V-13 presents Canberra-GRLVQ results where the best performance is seen when relevance rate is less than the learning rate. Finally, Figure V-14 presents classification results for Cosine-GRLVQ wherein performance is consistent with Figure V-11 with the best performance only achieved when the relevance rate is less than the learning rate and valued 0.0001 or 0.001.

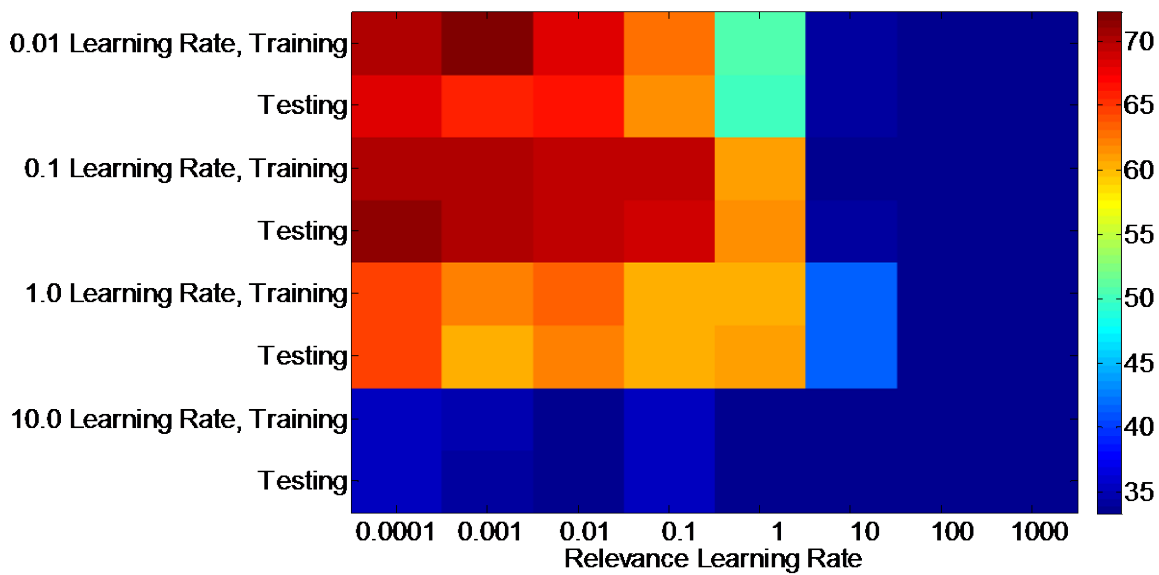


Figure V-12: Learning Rate vs Relevance Learning Rate for Squared Euclidean GRLVQ.

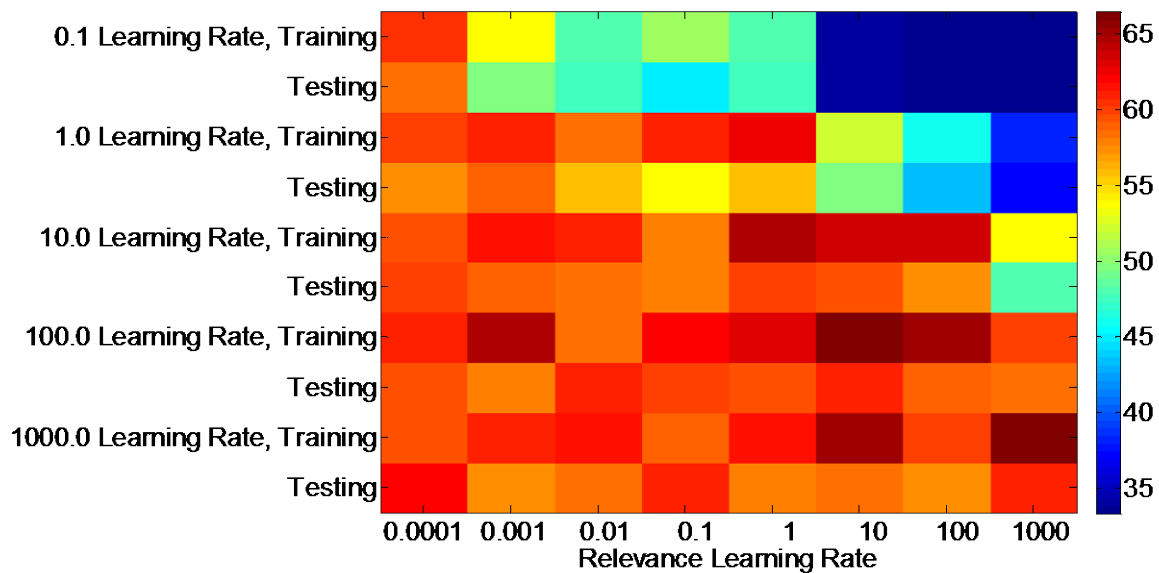


Figure V-13: Learning Rate vs Relevance Learning Rate for Canberra GRLVQ.

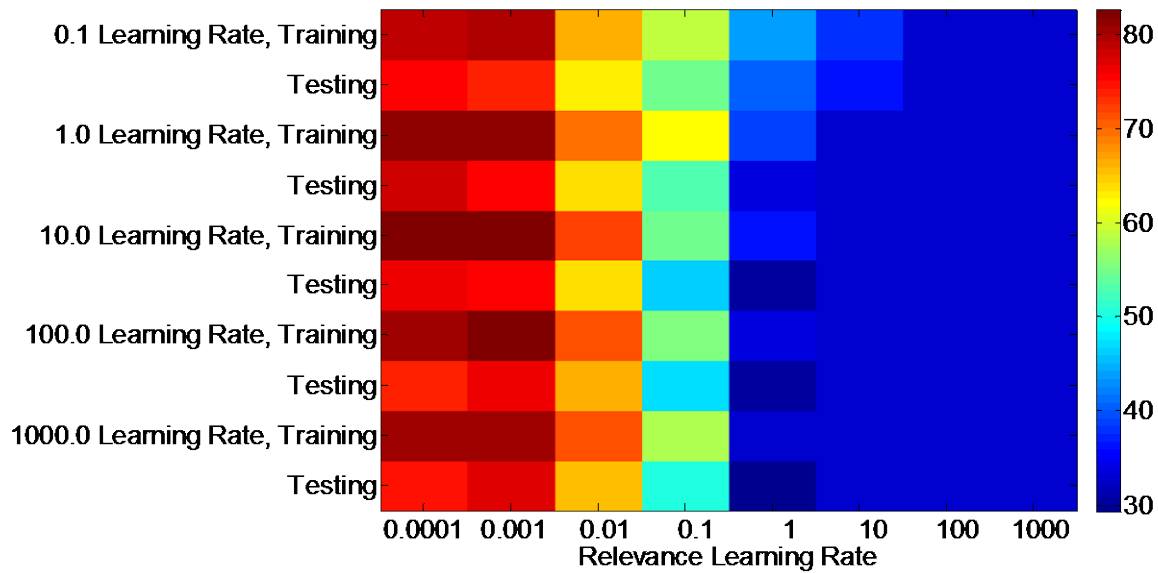


Figure V-14: Learning Rate vs Relevance Learning Rate for Cosine GRLVQ.

5.3.5.2 Distance Measure Extensions to GRLVQI

The extension of GRLVQ to GRLVQI involves components unrelated to the distance measure, PV gradient descent update or relevance gradient descent update. Therefore, algorithmically, the Cosine and Canberra versions of GRLVQ were extended to GRLVQI by incorporating the improvements of Section 5.2.2.

5.4 GRVLQI-D Extension for RF-DNA Fingerprinting

To extend the discussions in Sections 5.3.3–5.3.5 to GRLVQI for RF-DNA problems, a few general aspects must be considered: 1) LVQ architecture selection and 2) the interaction of GRLVQI factors of learning, relevance and conscience rates, LVQ architecture with the resultant classification and verification performance. For LVQ architecture selection, we will develop heuristics to determine the number of PVs to

instantiate and then consider the general impact of the number of PVs on Squared Euclidean GRLVQI classification and verification performance. To understand the interaction of these GRLVQI factors with performance, a full factorial ANOVA experiment will be considered (using Z-Wave data) with response surface methods used to find optimal settings. The algorithmic optimization approach is of particular interest for the Cosine and Canberra GRLVQI algorithms since there are no prior implementations of these from which to find reasonable settings.

5.4.1 LVQ Architecture Selection and Specification

As noted in Section 3.3.1.8, the literature is largely silent on the appropriate number of PVs, learning rates, PV initialization process except that one should use as many as possible [262] and that one needs at least one PV per class [299]. However, as seen in Schneider et al. [298], overfitting can occur in LVQ if too many PVs are instantiated. Additionally, since each PV must be moved in an iterative fashion, computation times necessarily increase when more PVs are considered. Therefore one should endeavor to instantiate a quantity of PVs that achieves good accuracy, avoids overfitting, and is not computationally expensive.

LVQ overfitting issues appear similar to overfitting problems in ANNs, as mentioned in [493], could suffer from similar problems as well since it is also a neural learning algorithm. An example of the overfitting effect is presented in Table V-2 which shows that an increasing number of ANN hidden nodes causes an increasing in training accuracy, but the resulting testing set accuracy does not similarly increase and reaches a

peak. LVQ architecture has similarities to ANNs and hence appropriately specifying the number of PVs could be critical to general LVQ performance. While the number of nodes is frequently determined empirically, e.g. [494–498], approaches for ANN architecture development exist and could be beneficial to LVQ algorithm performance.

Table V-2: Example of ANN Architecture Effects on ANN Performance, reproduced from [493].

INPUT NODES	HIDDEN NODES	OUTPUT NODES	TRAINING ACCURACY (%)	TESTING ACCURACY (%)
16	10	8	84.5	58.5
16	13	8	89.2	65.9
16	15	8	93.5	73.2
16	18	8	93.7	70.7
16	20	8	99.5	73.2
16	25	8	100	58.5

LVQ methods are considered to be generally robust to overfitting, as noted by Biehl et al. [470] and attributed to the Hebbian learning results. However, Schneider et al. [298] noted and presented results showing that LVQ can overfit on some datasets. Therefore consideration into the appropriate number of PVs is important. To illustrate the possibilities of LVQ overfitting, an example will be used. While the data examined by Clark [493] is not available, other academic datasets are. For this the small dataset Insects will be used; this dataset is from [499, 500] and consists of 3 data features, 3 classes, and 10 observations per class with no missing values. To examine potential overfitting effects, one randomly selected observations from each class was sequestered in a test set and an LVQ network was trained with the remaining 27 observations. The

number of PVs per class was then increased from 1-9, with a constant learning rate of $\epsilon(t)=0.1$ used throughout, 600 randomly generated training iterations were used. Mean test and training accuracy was then recorded for 100 replications. Table V-3 presents the results and shows that LVQ can be susceptible to overfitting and that robustness to overfitting is not universal for all LVQ algorithms in all applications.

Table V-3: Example of PV Architecture Effects on LVQ Performance on Insects.

NUMBER OF INPUT NODES (FEATURES)	PROTOTYPE VECTORS (PVs) PER CLASS	TRAINING ACCURACY (%)	TESTING ACCURACY (%)	MEAN COMPUTATION TIME (S)
3	1	68.0	69.67	0.34
3	2	80.7	73.7	0.50
3	3	84.4	77.3	0.62
3	4	85.9	69.7	0.69
3	5	89.0	70.0	0.86
3	6	89.5	68.0	1.16
3	7	89.3	66.0	1.09
3	8	89.9	66.7	1.59
3	9	91.5	68.0	1.59

Beyond employing as many PVs as possible, as suggested by [262], which can obviously lead to overfitting as shown in Table V-3, the LVQ field is largely bereft of literature on the number of PVs to initialize. However, the ANN field is replete with literature regarding appropriately selecting the number of hidden nodes in model development and includes heuristic approaches [304, 501] and algorithmic approaches

[502–504]. Of interest is if neural network heuristics for the number of hidden nodes can be extended to specifying the number of LVQ PVs.

5.4.1.1 Extending ANN Architecture Heuristics to LVQ

Lv et al. [253] considered 1 PV per class; for RF-DNA, Reising [51] used 10 PVs per class; however, for hyperspectral target detection, Mendenhall [244] used 5 PVs per class. While 1 PV per class is a minimum requirement for LVQ algorithms [299], and permits initializing each PV to the centroid (arithmetic mean) of its respective group as an easy and logical solution to the initialization problem, using too few PVs can yield poor results as empirically demonstrated in the academic example in Table V-3.

Although Mendenhall [244] mentioned using heuristics to determine the number of PVs for GRLVQI, they were not formalized for the family of LVQ algorithms considered. However, Gage [304] investigated and developed ANN architecture approaches where the size of the hidden layer was dependent on the number of inputs, number of exemplars, hidden layer weights, and/or the number of neurons at each layer. Although LVQ algorithms are ANNs, a few difficulties exist in extending general ANN methods to LVQ: firstly, the general LVQ architecture is not identical to ANN architecture, as described in Section 3.2; secondly, LVQ requires PVs to be designated to a class; and finally, LVQ does not have output nodes as seen in an ANN. Despite these differences, some empirical formulas for ANN architecture specification could be applicable to LVQ architecture specification.

Many heuristics considered by Gage [304] involve using the number of input features, N_F , the number of exemplars, N_{obs} , and the number of output layer nodes, N_{out} . Extending this to LVQ would see K being the number of input features and M representing the number of PVs; since LVQ does not have an output layer, one could interpret N_{out} as being either: A) nothing since LVQ does not have an output layer, in which case N_{out} would be treated as a constant 1 (thus N_{out} is equivalent to N_{PV} since N_{PV} is effectively the output layer in LVQ models), or B) we could logically view N_{out} as the number of classes, consistent with [274].

Basic neural network heuristics include the general following advice, that

$$N_{PV,Looney1} = aN_c \quad (5.15)$$

where a is a constant and N_c classes [250, p. 101]. While this is certainly suitable for LVQ architectures due to their underlying assumptions, it is not helpful in determining N_{PV} , and only provides the obvious lower bound of $N_{PV} = c$ for $a = 1$. However, an extension of this approach is seen in

$$N_{PV,WC} = aN_F \quad (5.16)$$

where a is used as a fraction [505]. In this form, a has variously been recommended as either 0.75 [506, 507] or 0.50 [508].

Looney [250, p. 91] presented another general heuristic of

$$N_{PV,Looney2} = \log_2 N_c \quad (5.17)$$

where N_c are the number of classes in the dataset, since this quantity will yield $N_{PV,Looney2} < N_c$ PVs it is not appropriate for LVQ models. Similar is the empirically

determined approach of Gorman and Sejnowski [494], noted as an effective heuristic for ANNs [509],

$$N_{PV,Gorman} = \log_2 T \quad (5.18)$$

where T is the number of input training patterns, however this terminology can be interpreted variously (depending on what one means by “pattern”) as either

$$N_{PV,Gorman1} = \log_2 N_F \quad (5.19)$$

or

$$N_{PV,Gorman2} = \log_2 N_{obs} \quad (5.20)$$

or possibly

$$N_{PV,Gorman3} = \log_2 (N_c \cdot N_{obs}). \quad (5.21)$$

Additional heuristics include one from Hayashi [250, p. 316; 510],

$$N_{PV,Hay} = q\sqrt{N_{out} \cdot N_F} \quad (5.22)$$

where q is a multiplier constant, set to 1 herein. Walczak and Cerpa [505] presented a heuristic based on [496, 511] that

$$N_{PV,Kur} = 2N_F + 1. \quad (5.23)$$

Gao et al. [501] presented the following heuristic,

$$N_{PV,Gao} = \sqrt{N_{out} \cdot N_F} + q, \quad (5.24)$$

with q being a constant between 1 and 10 and attributed it to [503]. Daqi and Shouyi [512] present the following heuristic

$$N_{PV,Daqi} = \sqrt{(N_{out} + 2) \cdot N_F} + 1, \quad (5.25)$$

Gage [304] presents a heuristic termed “Cover’s theorem”

$$N_{PV,Gage} < \frac{0.5N_{obs} - 1}{N_F + 1}. \quad (5.26)$$

which considered the number of exemplars, P , and data features [304, 501].

5.4.1.2 Developing LVQ Architecture Heuristics

Considering the heuristics in Section 5.4.1.1, the GRLVQI settings of [48, 247] and the absolute minimum of $N_{PV} = 1$ for the ZigBee RF-DNA data under analysis ($N_C = 4$, $N_{Feats} = 729$, $N_{obs} = 1500$), one arrives at Table V-4. Results for both $N_{out} = 1$ and $N_{out} = N_C$ are computed.

Table V-4: #PVs for RF-DNA Using Various Heuristics for ZigBee Data.

ORIGINATION	HEURISTIC	N_{PV}		
		$N_{out} = 1$ (N_{out} IGNORED)	$N_{out} = N_C$	INTERPRETED N_{PV}/N_C
ANNs	$N_{PV, Kur}$	*	*	1459†
	$N_{PV, Looney1}$	*	*	4
	$N_{PV, Looney2}$	*	*	2
	$N_{PV, Gage}$	*	*	1
	$N_{PV, Gorman1}$	*	*	10
	$N_{PV, Gorman2}$	*	*	11
	$N_{PV, Gorman3}$	*	*	13
	$N_{PV, Gao1}$	28	55	14, 28
	$N_{PV, Hay}$	27	54	14, 27
	$N_{PV, Daqi}$	48	68	17, 48
	$N_{PV, WC}$	*	*	365-547†
LVQ	$N_{PV, Min}$	*		1
	$N_{PV, Mendenhall}$	*		5
	$N_{PV, Reising}$	*		10

*indicate heuristic is not a function of N_{out} and hence this quantity is not computed

† indicates obviously unreasonable values for N_{PV}

Based on the results presented in Reising [51] and Dubendorfer [91], both for $N_{PV} = 10$ per class for RF-DNA Fingerprints, we can safely exclude the number of PVs suggested by $N_{PV, WC}$ and $N_{PV, Kur}$ as considerably too many. However, the remaining heuristics suggest numbers of PVs that appear reasonable.

A vector of quantities of PVs (per class) to consider was formulated as:

$$N_{PV} = [1, 5, 7, 8, 9, 10, 11, 12, 13, 15, 20, 27, 37, 48]. \quad (5.27)$$

$N_{PV} = [7, 8, 9, 11, 12, 13]$ per class were also considered in order to search for suitable operating points across the heuristic space and around the nominal setting of 10 N_{PV}/N_C . Values of 14 and 17 N_{PV}/N_C were not considered since these are close to 15 N_{PV}/N_C to avoid superfluous computational runs. Values above 48 N_{PV}/N_C were not initially considered due to the extra computation time required, and thus these would only be considered if the results indicate a potential utility in exploring these settings.

Figure V-16 considers GRLVQI results on the ZigBee dataset at 14 dB for the N_{PV} values in (5.27). The preliminary results in Figure V-16 shows that overfitting would be an issue if too many PVs were instantiated, $N_{PV} > 20$, and that poor accuracy would result if too few PVs were instantiated, $N_{PV} < 9$. From Figure V-16, $N_{PV} = 13$ offers overall higher training, testing, and validation accuracy than $N_{PV} = 10$; additionally, the overall difference between higher training, testing, and validation accuracy are small when compared to $N_{PV} \geq 15$.

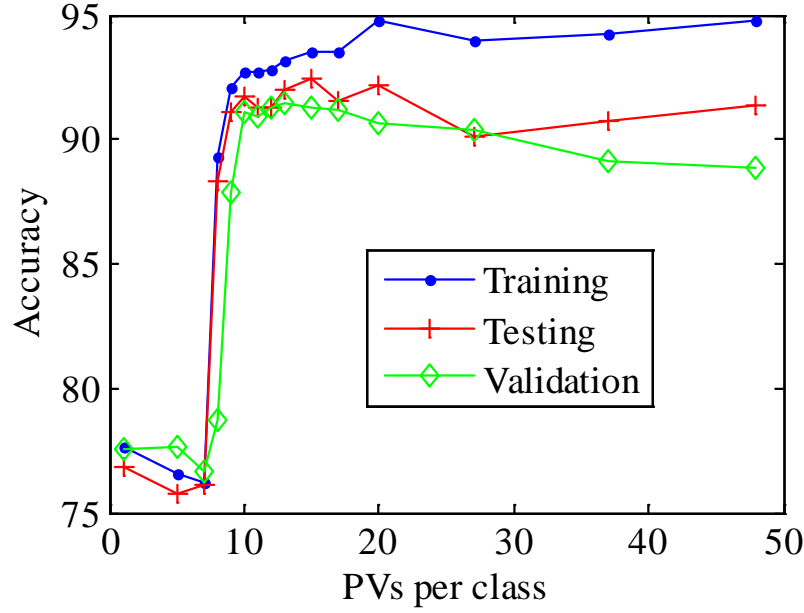


Figure V-15: GRLVQI Classification Results on ZigBee RF-DNA Fingerprints at 14 dB Using Various PVs/class.

Figure V-16 presents classification performance results from considering Squared Euclidean GRLVQI for $N_{PV} = [10, 13]$ with the ZigBee RF-DNA Fingerprints. As seen in Figure V-16 classification performance appears comparable for $\text{SNR} \geq 10\text{dB}$, with $N_{PV} = 13$ offering a slight improvement in gain of +0.41dB (training) and +0.51dB (testing) at 90% accuracy. However classification performance appears markedly improved for low SNR, and between 5dB and 10dB GRLVQI with $N_{PV} = 13$ offers a gain of +1.85dB (training) and +2.27dB (testing) at 70% accuracy.

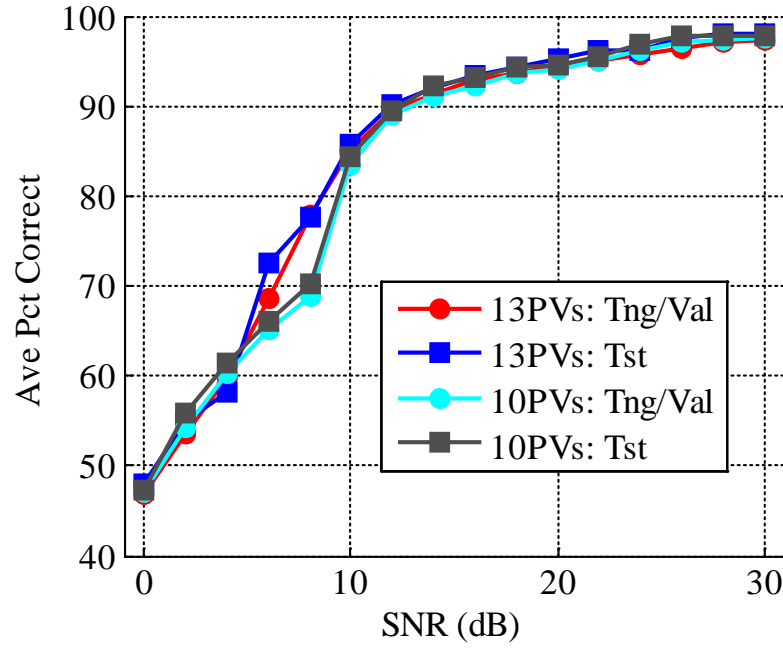
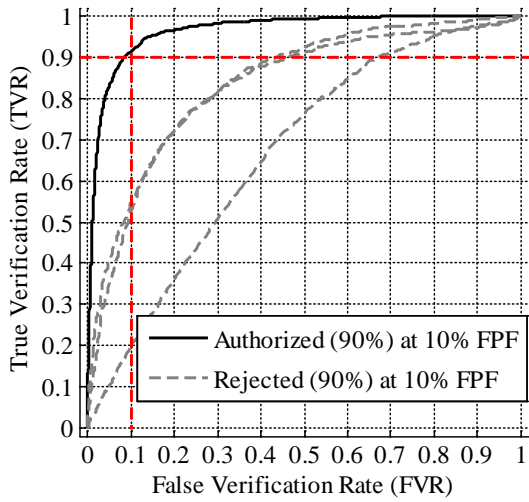
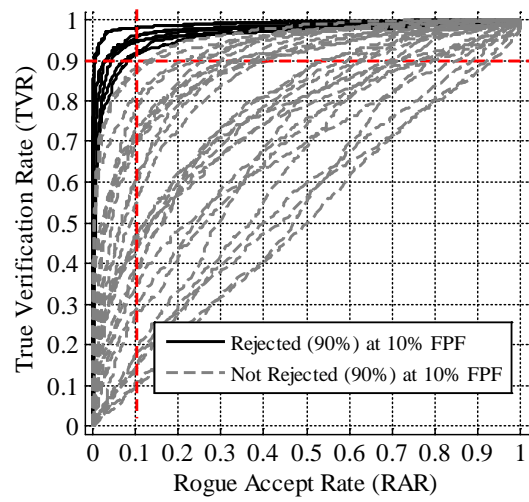


Figure V-16: GRLVQI Classification Performance with 10 PVs versus 13 PVs.

When considering verification accuracy with Squared Euclidean GRLVQI using $N_{PV} = 13$, one can see in Figure V-17 to Figure V-19 that more structure is seen when compared to the verification results seen in Section III for $N_{PV} = 10$.

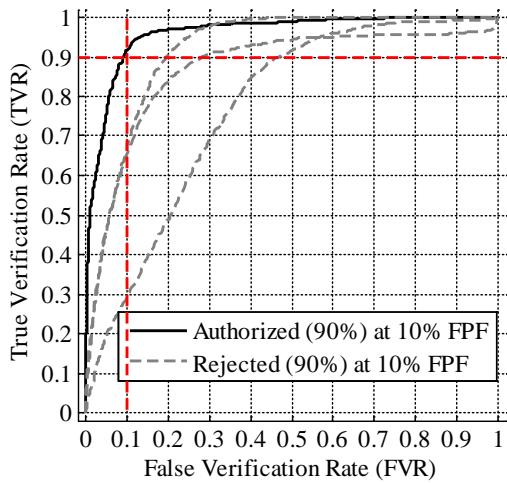


a) Authorized

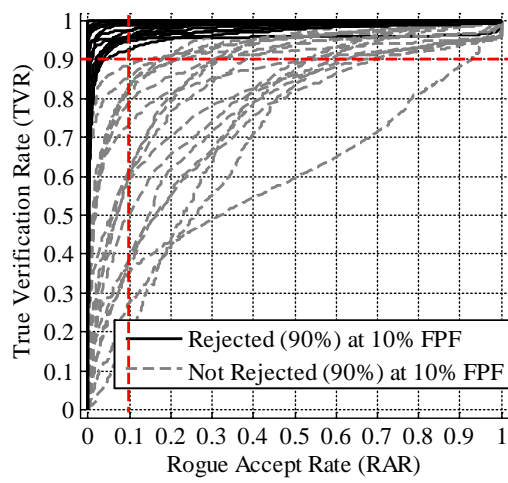


b) Rogue

Figure V-17: Verification Performance in GRLVQI with 13 PVs at 8dB.

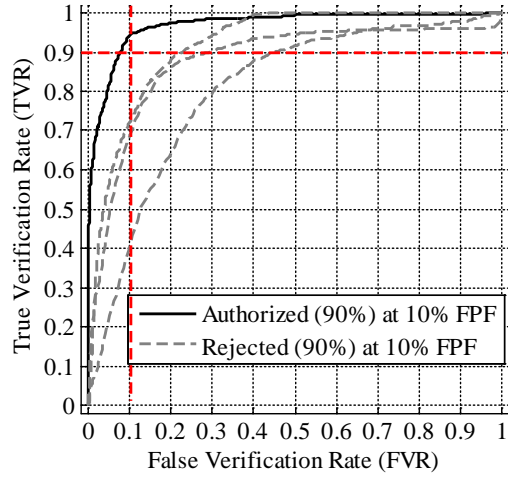


a) Authorized

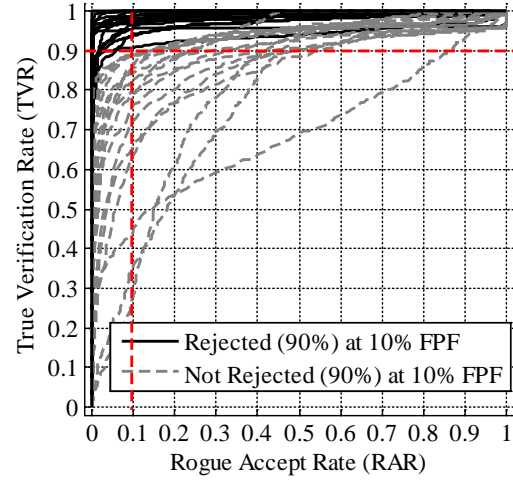


b) Rogue

Figure V-18: Verification Performance in GRLVQI with 13 PVs at 14dB.



a) Authorized



b) Rogue

Figure V-19: Verification Performance in GRLVQI with 13 PVs at 18dB.

Table V-5 presents an overall comparison of classification and verification performance for Squared Euclidean GRLVQI with $N_{PV} = [10, 13]$. Overall, classification performance is largely improved with 13 PVs while verification performance is greatly improved for low SNR and slightly worse for higher SNR. Overall, one can conclude that 13 PVs offers measurable performance improvements over the 10 PVs. However, possible changes to learning, relevance and conscience rates have not been considered.

Table V-5: Relationship between PVs and Classification/Verification Performance

N_{PV}	Classification Performance				Verification Performance					
	SNR (dB) at 90% C		$AUCC_{TNG}$	$AUCC_{TST}$	% Authorized or % Rogue Rejected at 8dB		% Authorized or % Rogue Rejected at 14dB		% Authorized or % Rogue Rejected at 18dB	
	TNG	TST			Authorized	Rogue	Authorized	Rogue	Authorized	Rogue
10	12.92	12.39	24.99	25.24	0%	0%	25%	47.22%	25%	63.88%
13	12.51	11.88	25.27	25.51	25%	22.22%	25%	50%	25%	52.78%

5.4.2 Experimental Design for GRLVQI-D Algorithmic Settings

Employing experimental designs to find optimal algorithm settings has been seen in hyperspectral anomaly detection research, c.f. [513–520], but not in prior RF-INT efforts. However, herein, determining appropriate algorithmic settings is of prime interest since neither Cosine GRLVQI nor Canberra GRLVQI algorithms have been previously developed or applied to RF-DNA problems. Therefore it is unknown what settings are appropriate for these algorithms.

Following the discussions in Sections 5.3.3–5.3.5, a few observations can be made, 1) that Cosine and Squared Euclidean variants of LVQ, RLVQ, GLVQ and GRLVQ perform similarly well in classification of Fisher Iris; 2) that Cosine LVQ variants perform best with both a learning rate 10 times or greater and a relevance rate 1/10 to 1/100 of that seen in Squared Euclidean LVQ variants; and 3) that Canberra variants similarly performed best with both a learning rate 10 times or greater than Squared Euclidean, but appeared invariant to relevance learning rate. Additionally, in Section 5.4.1, we learned that changing the number of PVs can significantly impact GRLVQI performance.

5.4.2.1 Full Factorial Model

To determine optimal settings for Squared Euclidean GRLVQI, Cosine GRLVQI, and Canberra GRLVQI algorithms, a full factorial experiment was considered. Table V-6 presents the 3^5 design wherein the middle (0) design settings are those employed by Reising [51], the high and low settings for learning and relevance rates are magnitudes of 10 above and below, respectively, the middle settings per the observations in Sections

5.3.3–5.3.5. Two conscience rates are present in GRLVQI and the scale of these differs from the learning and relevance rates; Table III-3 presented training steps and corresponding explored conscience rates where γ is seen to be initialized as 2.0 and reach an absolute minimum (after many training steps) of 0.75, and β is initialized 0.35 and reach an absolute minimum of 0.10. To account for this range and explore other possible good settings, the full factorial experiment explores a low setting of 0.5 and a high setting of 4.5 for γ and a low setting of 0.15 and a high setting of 0.55 for β . Additionally, the number of PVs is considered as a fifth factor where 13 PVs per class is considered as the high value and 7 PVs per class is considered as the low value, per the discussion in 5.4.1.

Table V-6: Experimental Design Region for GRLVQI.

	FACTORS				
	FACTOR A	FACTOR B	FACTOR C	FACTOR D	FACTOR E
FACTOR LEVEL	LEARNING RATE (ϵ)	RELEVANCE RATE (ξ)	CONSCIENCE RATE 1 (γ)	CONSCIENCE RATE 2 (β)	NPV
LOW (–)	0.0025	0.0005	0.5	0.15	7
MIDDLE (0)	0.025	0.005	2.0	0.35	10
HIGH (+)	0.25	0.05	4.5	0.55	13

Employing the settings from Table V-6 yields a total of 243 different setting combinations per GRLVQI-D variant. To consider all of these possible operating points, Z-Wave RF-DNA data, as described in Section III and employed in [49], was used due to the much smaller size of this data set and its signal similarity to ZigBee. Appendix K presents mean training and testing AUCC along with mean verification AUC values experimental results from Z-Wave for the Cosine, Canberra and Squared Euclidean

GRLVQI algorithms grouped by distance measures for the experimental design in Table V-6. To expedite the computational process, the baseline Squared Euclidean GRLVQI algorithm employed MATLAB compiled c-code (*.mex) files, which were compiled via the approach in Appendix L.

5.4.2.2 Response Surface Methodology

After the experimental runs in Section 5.4.2.1 were complete, a second order model with squared terms and two-way interactions was considered:

$$J(x) = B_0 + \sum_{i=1}^s B_i x_i + \sum_{i,j,i=1}^s B_{i,j} x_i x_j + \sum_{i,j,i=1}^s B_{i,i} x_i^2. \quad (5.28)$$

where s represents the number of factors, B terms are coefficients solved for via a general linear model, and x represents a given factor [513]. Two initial second order models were created per algorithm with all parameters and interactions (termed “Full Model”) after applying (5.28) with either classification (mean AUCC) or verification (mean AUC) accuracy as the dependent variable. All models were statistically significant using $\alpha = 0.05$, but not all features and interactions were significant, reduced models were therefore created by creating a second order model that only contained main effects (factors in Table V-6, whether or not significant) and any significant second order effect. Table V-7 presents an overview of the second order models by reporting R^2 and adjusted R^2 values for both the full and reduced models.

As seen in Table V-7, the classification models from Squared Euclidean data explains a significant amount of variance in the data while the verification based models

do not explain much variation. When considering the Cosine GRLVQI models, both the classification and verification models explain most of the variation in the data; however neither the classification nor the verification based models explain much variation when considering the Canberra GRLVQI results.

Table V-7: Overview of Second Order Models.

		ALGORITHM					
		SQUARED EUCLIDEAN GRLVQI		COSINE GRLVQI-D		CANBERRA GRLVQI-D	
	DEPENDENT VARIABLE	CLASS.	VER.	CLASS.	VER.	CLASS.	VER.
FULL MODEL	R^2	0.900	0.246	0.942	0.829	0.259	0.408
	R^2 ADJUSTED	0.891	0.178	0.937	0.814	0.193	0.355
REDUCED MODEL	R^2	0.898	0.241	0.938	0.824	0.215	0.399
	R^2 ADJUSTED	0.892	0.195	0.936	0.817	0.188	0.378

Table V-8 presents variables that were deemed statistically significant in the full model. Again, in all reduced models main effects were included for completeness. In Table V-8, an “X” indicates that a variable is statistically significant, at $\alpha = 0.05$, while a “?” indicates that a variable has a p -value between 0.05 and 0.10, which should be considered as statistically significant at $\alpha = 0.05$, per [369]. As seen in Table V-7, the R^2 and adjusted R^2 are largely unchanged when considering the reduced models, indicating that the removed features were not explaining much variation in the data.

Table V-8: Features Significant Per Model.

FEATURE	MODEL					
	SQUARED EUCLIDEAN GRLVQI		COSINE GRLVQI-D		CANBERRA GRLVQI-D	
	CLASS.	VER.	CLASS.	VER.	CLASS.	VER.
ϵ	X	X	X	X		X
ξ	X				X	
γ	X					
β	X					X
N_{PV}	X	X	X	X	X	X
ϵ^2	X	X	X	X		X
ξ^2	X				X	
γ^2	X	X				
β^2	X					
N_{PV}^2	X					
$\epsilon \times \xi$	X	X			X	X
$\epsilon \times \gamma$	X		?			
$\epsilon \times \beta$						
$\epsilon \times N_{PV}$	X		X			
$\xi \times \gamma$						
$\xi \times \beta$						
$\xi \times N_{PV}$						
$\gamma \times \beta$						
$\gamma \times N_{PV}$?		
$\beta \times N_{PV}$	X				X	

5.4.2.3 Setting Optimization

As mentioned in Chapter 3, determining appropriate settings for LVQ algorithms is a largely untouched domain; however, after finding reduced second order models, one can solve for optimal algorithmic settings where the target are the dependent variables (either mean classification or mean verification accuracy). Determining appropriate settings is of critical important for distance measure variants of GRLVQI since these have unknown operating characteristics.

Constrained nonlinear optimization, or interior point optimization, consistent with [521–523] was used to maximize the final, reduced, second order models. A constrained minimization (where the target accuracies were negated since the goal of maximization is possible by minimizing a negation) was considered where a finite-difference approximation was computed by starting with an initial estimate (the baseline GRLVQI settings). The minimization was constrained between the minimum and maximum values seen in Table V-6 to avoid computing values outside those explored (e.g. when considered unbounded optimization yielded settings far outside the design space, with magnitudes ranging from 10^{13} to 10^{42}). The optimal solution was then computed via sequential quadratic programming (SQP) [524, 525] wherein a line search was employed, consistent with [524–526].

Resultant optimal algorithmic settings for each factor are presented in Table V-9. Here, settings are grouped in pairs of rows by algorithm and then by whether mean classification AUCC or mean verification AUC were used at the target. Evident in Table V-9 is that only $N_{PV} = 7$ was consistently found as optimal between algorithms.

Otherwise, most factors had different optimal algorithmic settings. Additionally, all optimal settings were different from the baseline settings as employed by [51].

Table V-9: Optimized Algorithms Settings for Z-Wave Data.

		FACTORS				
		FACTOR A	FACTOR B	FACTOR C	FACTOR D	FACTOR E
ALGORITHM		LEARNING RATE (ϵ)	RELEVANCE RATE (ξ)	CONSCIENCE RATE 1 (γ)	CONSCIENCE RATE 2 (β)	N_{PV}
SQUARED EUCLIDEAN GRLVQI	CLASS.	0.1497	0.0005	4.5	0.3128	7
	VER.	0.1481	0.05	0.5	0.15	7
COSINE GRLVQI-D	CLASS.	0.1376	0.05	4.5	0.55	7
	VER.	0.135	0.0005	0.5016	0.15	7
CANBERRA GRLVQI-D	CLASS.	0.25	0.032	0.5	0.15	7
	VER.	0.25	0.032	0.5	0.15	7
BASELINE	--	0.025	0.005	2.0	0.35	10

5.4.3 GRLVQI-D Performance Results

Classification and verification performance can be considered using the optimized algorithmic settings. Z-Wave classification performance will be considered relative to the baseline classifier settings of Reising [51]. Three sets of classification results are considered in Figure V-20 through Figure V-24. Figure V-20 presents training (TNG) and testing (TST) classification results from the baseline Squared Euclidean GRLVQI algorithm, the Squared Euclidean GRLVQI algorithm using the Classification-based optimized settings in Table V-9, and the Squared Euclidean GRLVQI algorithm using the Verification-based optimized settings in Table V-9. Noticeably, classification performance appears markedly improved when using either optimized setting, which also

have fewer PVs, $N_{PV} = 7$, and thus are computationally simpler algorithms. The Classification-based optimized Squared Euclidean GRLVQI shows an improvement in gain of +1.98 dB (TNG) and +1.94 dB (TST) at 90% accuracy; the Verification-based optimized Squared Euclidean GRLVQI shows an improvement in gain of +1.31 dB (TNG) and +1.48 dB (TST).

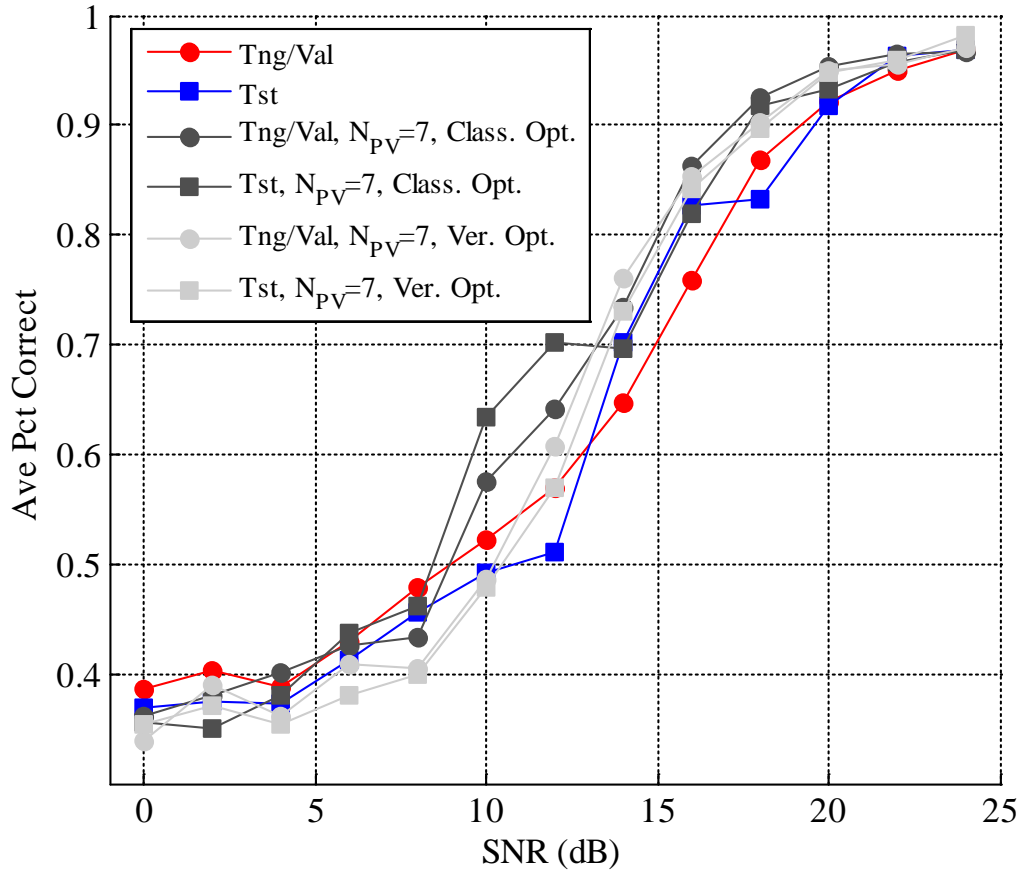
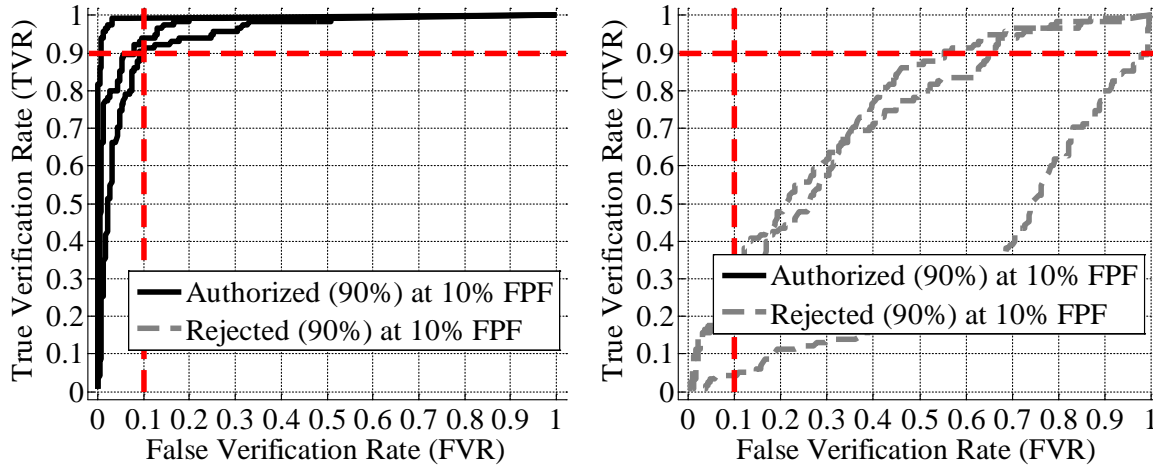


Figure V-20: GRLVQI Classification Performance Using Squared Euclidean Distance Using Optimized Algorithmic Settings.

Figure V-21 presents the verification accuracy of both optimized Squared Euclidean GRLVQI algorithms; one can see that the Classification-based Squared

Euclidean GRLVQI Figure V-21a offer 100% verification accuracy, which improves upon the 33.33% classification accuracy of the baseline, reported in Section III and [49]. Additionally, the mean verification AUC of 0.9707 is slightly higher than the mean AUC of the baseline, 0.9363. When considering the Verification-based Squared Euclidean GRLVQI performance in Figure V-21b, the performance is noticeably poor, with no devices authorized at 10% FVR and 90% TVR. Additionally, the curves in Figure V-21b are significantly worse than baseline with a mean AUC of 0.5916. Overall, it's evident that the optimized settings from the Classification-based Squared Euclidean GRLVQI offer improved performance over baseline, while using the optimized settings from the Verification-based Squared Euclidean GRLVQI classifier offers comparably unreasonable verification performance.



a) Classification-Based Optimization

b) Verification-Based Optimization

Figure V-21: GRLVQI ID Verification Performance in Squared Euclidean GRLVQI using Optimization Settings at 20dB for Z-Wave Dataset.

Classification results from the Canberra GRLVQI-D classifier are presented in Figure V-22 for the Classification-based and Verification-based optimized settings with the Z-Wave data. The performance of both is dramatically below the baseline Squared Euclidean GRLVQI algorithm. Figure V-23 presents the verification accuracy of both optimized Cosine GRLVQI algorithms; one can see that neither the Classification-based Canberra GRLVQI-D in Figure V-23a nor the Verification-based Canberra GRLVQI-D in Figure V-21b perform well. Additionally, the curves in Figure V-21b are significantly worse than baseline with a mean AUC of 0.5916. Overall, it's evident that Canberra GRVLQI-D, at least with the considered settings, appears unsuitable for RF-DNA applications.

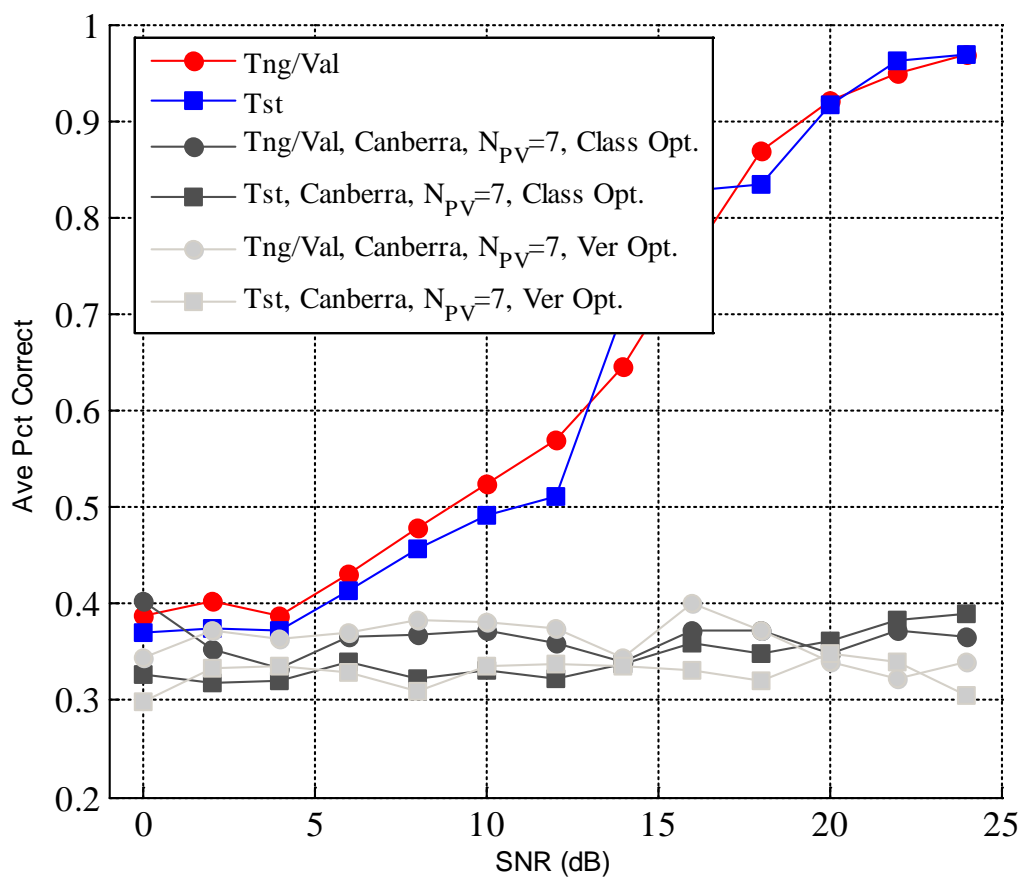


Figure V-22: GRLVQI-D Classification Performance Using Canberra Distance Using Optimized Algorithmic Settings.

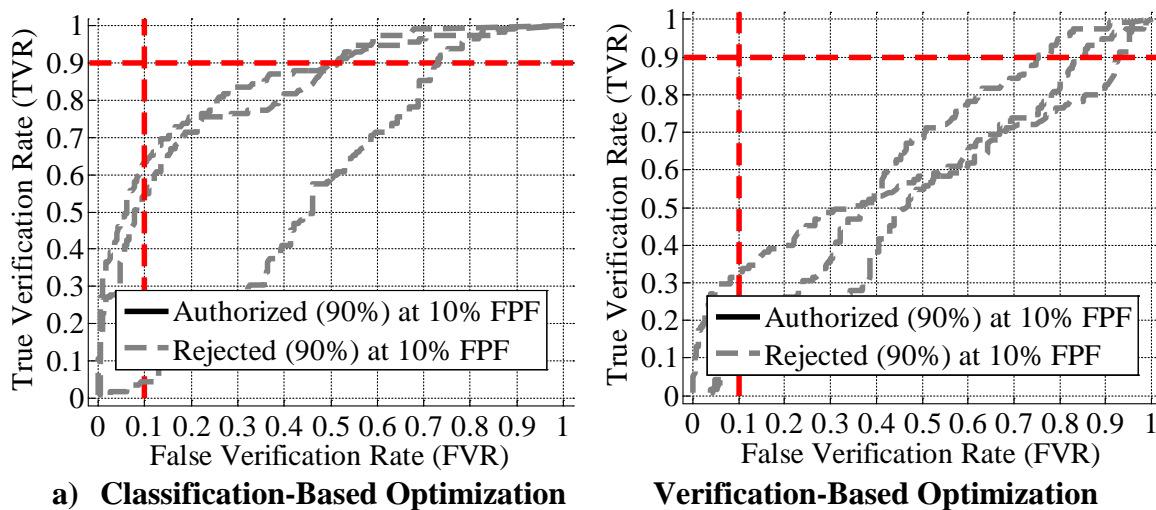


Figure V-23: GRLVQI-D ID Verification Performance in Canberra GRLVQI using Optimization Settings at 20dB for Z-Wave Dataset.

Classification results from the Cosine GRLVQI-D classifier are presented in Figure V-24 for the Classification-based and Verification-based optimized settings with the Z-Wave data. In contrast to the Canberra GRLVQI-D algorithms of Figure V-22, the Cosine GRLVQI-D classification results offer improved performance over the baseline Squared Euclidean GRLVQI algorithm. The Classification-based optimized Squared Euclidean GRLVQI shows an improvement in gain of +1.57 dB (TNG) and +1.91 dB (TST) at 90% accuracy; the Verification-based optimized Squared Euclidean GRLVQI shows an improvement in gain of +1.67 dB (TNG) and +1.84 dB (TST). Performance is thus similar to the optimized Squared Euclidean GRLVQI algorithm in Figure V-20.

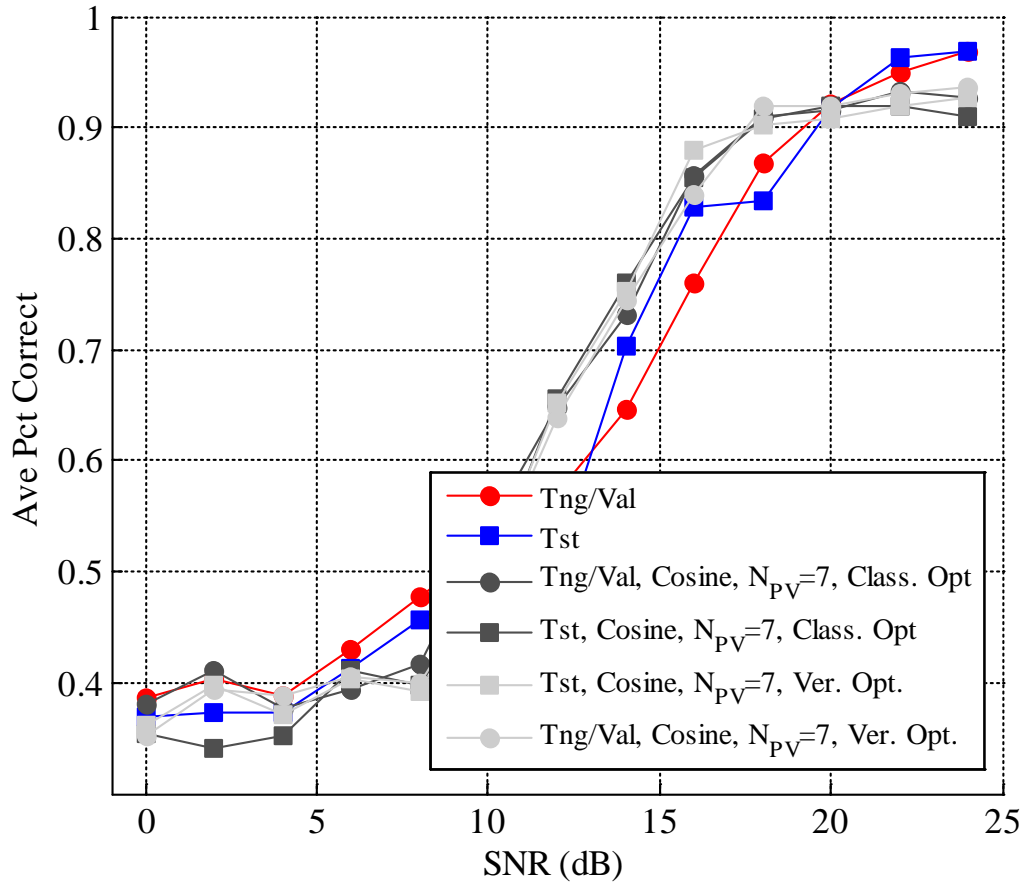


Figure V-24: GRLVQI Classification Performance Using Cosine Distance Using Optimized Algorithmic Settings.

Figure V-25 presents the verification accuracy of both optimized Cosine GRLVQI-D algorithms; one can see that both the Classification-optimized Cosine GRLVQI-D Figure V-25a and Verification-optimized Cosine GRLVQI-D offer 66.6% verification accuracy, which improves upon the 33.33% verification accuracy of the baseline, as reported in Section III and [49], but is slightly worse than the 100% verification accuracy of the Classification-optimized Squared Euclidean GRLVQI algorithm in Figure V-21a. However, the mean verification AUC of both Cosine

GRLVQI-D variants is 0.9712 which is equivalent to the mean verification AUC of 0.9707 for the Classification-optimized Squared Euclidean GRVLQI algorithm.

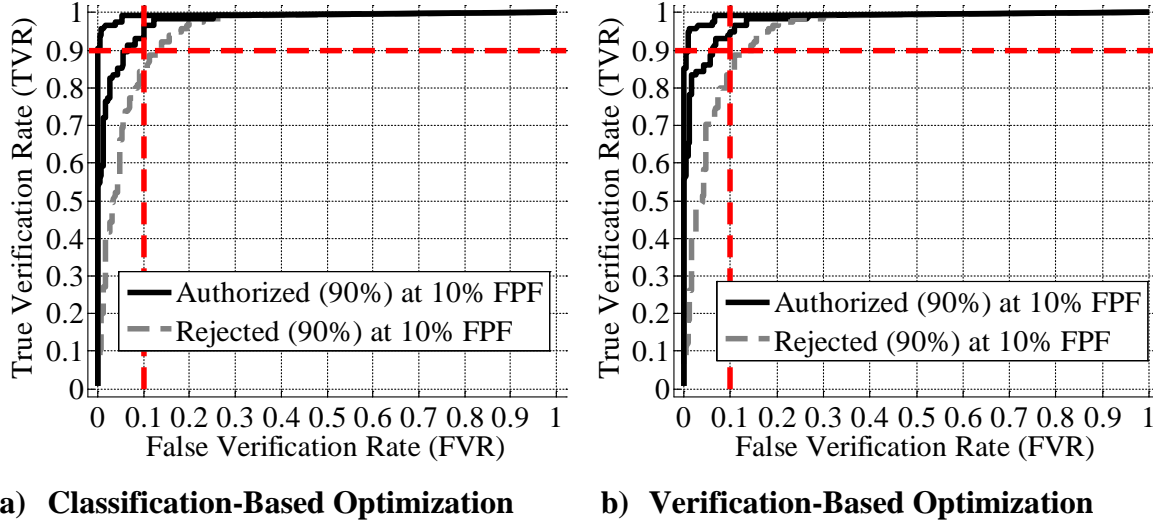


Figure V-25: GRLVQI ID Verification Performance in Cosine GRLVQI using Optimization Settings at 20dB for Z-Wave Dataset.

Table V-10 presents an overall comparison of classification and verification performance for the Squared Euclidean GRLVQI algorithm, the Cosine GRLVQI-D algorithm, and the Canberra GRVLQI-D algorithm. Overall, classification performance is largely improved over baseline when using either the Optimized (either Classification or Verification based) Cosine GRLVQI-D algorithm or the Classification-optimized Squared Euclidean GRLVQI algorithm. Canberra GRLVQI-D offers no performance benefits and thus it is not further considered.

Table V-10: Z-Wave Optimized Algorithms Results for Z-Wave Data.

ALGORITHM		RESULT					
		CLASSIFICATION				VERIFICATION AT 20 dB	
		RAP_{TNG}	RAP_{TST}	SNR Gain (dB) Relative to TST Baseline at 90%C		% Auth.	Mean AUC
				TNG	TST		
SQUARED EUCLIDEAN GRLVQI	NONE (BASELINE)	1.01	1.00	+0.4	0.00	33.33%	0.936
	CLASS.	1.06	1.06	+0.44	+1.94	100%	0.971
	VER.	1.03	1.01	+0.23	+1.48	0%	0.592
COSINE GRLVQI-D	CLASS.	1.03	1.01	+0.06	+1.91	66.67%	0.971
	VER.	1.02	1.03	+0.23	+1.84	66.67%	0.971
CANBERRA GRLVQI-D	CLASS.	0.58	0.54	N/A		0%	0.740
	VER.	0.58	0.53	N/A		0%	0.560

Appendix M presents an extension of the Z-Wave optimized GRLVQI and GRLVQI-D algorithms with ZigBee data. While the optimized algorithms improve performance for Z-Wave data, the results in Appendix M illustrate the difficulty in applying optimized settings from one dataset to a different dataset. Thus, if ZigBee devices are of specific interest, one would desire to optimize GRLVQI and GRLVQI-D algorithmic settings for these devices.

5.4.4 Results Interpretation

Overall, the process and methodology presented in this chapter enable one to create distance measure variants of LVQ algorithms including LVQ, RLVQ, GLVQ, GRLVQ, and GRLVQI. The derivative skeleton presented in Section 5.2.2.4 further

enables one to make any reasonable change to the cost function of LVQ family of algorithms. The N_{PV} heuristics and optimization scheme provide a further approach for selecting reasonable settings for these algorithms.

Optimization of GRLVQI was considered using both Classification accuracy and Verification accuracy as an objective. Z-Wave data was employed due to the smaller size of the dataset and the requirement for a multitude of algorithmic run, as seen in Appendix K. When optimized settings were considered and evaluated on Z-Wave data, both the Classification-optimized Squared Euclidean GRLVQI algorithm and the Classification and Verification-optimized Cosine GRLVQI-D algorithms offered better performance than the baseline settings of [51]. The results for both the optimized Squared Euclidean GRLVQI and optimized Cosine GRLVQI-D algorithms are reasonable and hence the optimization method and process show efficacy for finding robust points when other devices are under analysis, and for recommending new operating points for either new algorithms, such as Cosine GRLVQI-D, or new signal modalities.

VI. Improvements to the RF-DNA Fingerprinting Process

Adorn thyself with simplicity

–MARCUS AURELIUS, 121-180

In operation, as described in Chapter II, the AFIT RF-DNA process consists of two main elements, including signal collection (accomplished using various signal collection equipment) and post-collection processing (accomplished using software). After collection, the data is digitally filtered and processed to create samples at various desired analysis SNR levels. Subsequently, RF-DNA fingerprints are computed and various *device classification* schemes applied for model development. In computed RF-DNA fingerprints, as described in Section 2.4, the signal Region Of Interest (ROI) is divided into multiple subregions (N_R total), each with N_S time samples per subregion. In each subregion, mathematical moments of mean (μ), variance (σ^2), skewness (γ), and kurtosis (κ), using (2.9), (2.6)–(2.8) respectively, are computed to provide insight into the distribution shape about its mean. Of interest in this chapter are potential improvements that can be made to the RF-DNA Fingerprinting process by leveraging research and methods in statistical data analysis, and simulation studies.

6.1 Introduction

First, Section 6.2 will examine data analysis methods and possibly underlying reasons for the dominance of phase features in RF-DNA Dimensional Reduction Analysis (DRA). Then, Section 6.3 will consider extensions of Simulation, an

operations research tool for examining steady state conditions from a time sample application [136], to RF-DNA.

6.2 Normalization, Standardization and Phase Feature Dominance

Prior works, such as [113], have concluded that phase features were significantly more useful for classification and verification than either amplitude or frequency features. However, no reasons for this observation have been determined. Three possible reasons for this result are hereby posited: 1) the mean centering and maximum scaled normalization in [19] produces this as an artifact, 2) the signal modulation method, e.g. ZigBee is Phase modulated as described in Section 2.2.1, is reflected in this result, and 3) intrinsic qualities of amplitude, phase, and frequency responses are being represented. Of interest here are considering 1) and 3) since 2) requires collecting signals from a wide variety of devices.

6.2.1 Phenomenology of Amplitude, Frequency, and Phase

Conclusive reasons for the dominance of phase features in RF-DNA research do not exist; however, various potential reasons do exist and are related to the phenomenology of amplitude, frequency, and phase. Amplitude, frequency, and phase are related quantities that can describe a signal. All three quantities are inter-related via the expressions described in (2.2)–(2.4) and [64, 191, 192, 527]. In computation for a real-valued signal, instantaneous amplitude is computed as the magnitude at a given point in time, instantaneous phase is then computed as the angle of the signal's Hilbert

transform, and finally, instantaneous frequency is computed as the gradient of instantaneous phase [64, 191, 192, 527].

While environmental characteristics may be captured in all three measurements [528], they are more pronounced in amplitude, e.g. amplitude modulated (AM) radio signals are more susceptible to storm disturbances than frequency modulated (FM) radio signals [529]. The ZigBee and Z-Wave devices of interest herein are Phase Modulated (PM) signals; PM signals are designed so that amplitude variations are small with ideally constant amplitude [527]. Additionally, in RF-DNA research relatively narrow frequency regions are generally isolated through filtering such that the signal itself may not vary much in frequency. Additional reasons for phase features being most significant could include phase noise due to production variations [530] and that phase variations have a more irregular pattern, short settling duration and a smaller dynamic range [112]. Therefore, it seems reasonable that phase features dominate, and especially for the PM signals.

6.2.2 Normalization and Standardization

When one examines a boxplot of the ZigBee features, Figure VI-1, it is seen that phase, amplitude and frequency features have different distributions. Boxplots are akin to plotting a histogram in condensed form [531, 532], thus permitting the distribution of multiple features to be evaluated side-by-side. The boxplot format presented in Figure VI-1 employs a “compact” format with a black dot indicating the median, thick blue lines to show the range from the 25th to 75th percentiles, thin blue lines to encompass all other

non-outlier data, and small blue circles to indicate outliers [533]. Figure VI-1 shows that the distribution and medians of phase features are more constrained than amplitude or frequency features. Additionally, phase features appear to have fewer outliers.

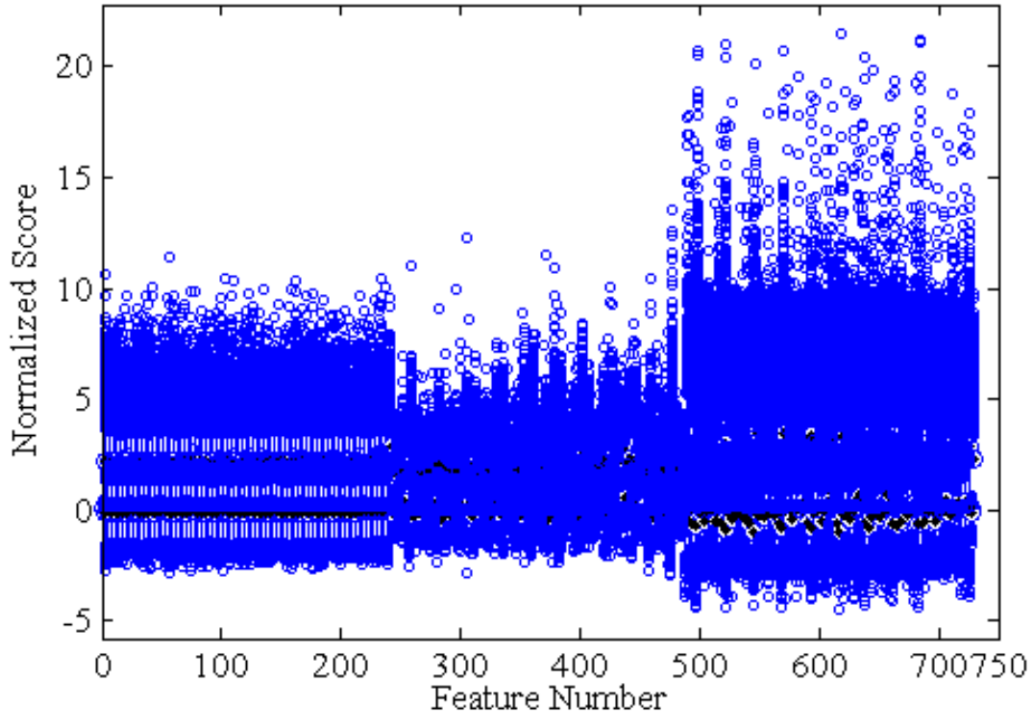


Figure VI-1: Boxplot of ZigBee RF-DNA Features at $SNR = 10dB$ for Authorized Devices Using the Nominal Mean Centering and Maximum Scaled Normalization process of [18, 19].

Due to characteristics of PM signals, any data normalization process could further impact feature relevance. The nominal RF-DNA Fingerprinting process incorporates a mean centering and maximum scaled normalization approach seen in (2.5) of Section 2.4. While mean centering and maximum scaled normalization does not appear in reviews of normalization methods, e.g. [534], this approach is consistent with various applications,

c.f. [535–541]. However, the reason for using this approach is rarely provided; one exception is Cobb et al. [18] who indicated that this normalization approach was used to account for any “uncontrolled power variation.”

Classifiers and neural network approaches frequently work best with input data normalized by some means [542]; however, it is very common to employ standard score normalization (standardization) --c.f. [330, 534, 543]. The boxplots in Figure VI-1 display that the data has different ranges for amplitude, frequency and phase, and hence examining any issue with the mean centering and maximum scaled normalization approach is important.

To examine the effect of normalization on RF-DNA, a revised RF-DNA normalization was therefore applied in the form of

$$\bar{g}_c[n] = \frac{g[n] - \mu_g}{std(g_c[n])}, \quad (6.1)$$

where g in (6.1) represents the signal of interest, per the respective RF-DNA fingerprint elements in (2.2)–(2.4) for $n = 1, 2, \dots, N_S$, where N_S represents the number of samples in the region of interest (ROI), and μ_g represents the mean of the g -th fingerprint element.

After standardizing the data, the RF-DNA fingerprinting process was followed, otherwise unaltered, and the resultant standardized RF-DNA features are presented in Figure VI-2. The data means now appear more centered and the ranges of the distributions of amplitude and frequency appear more constrained.

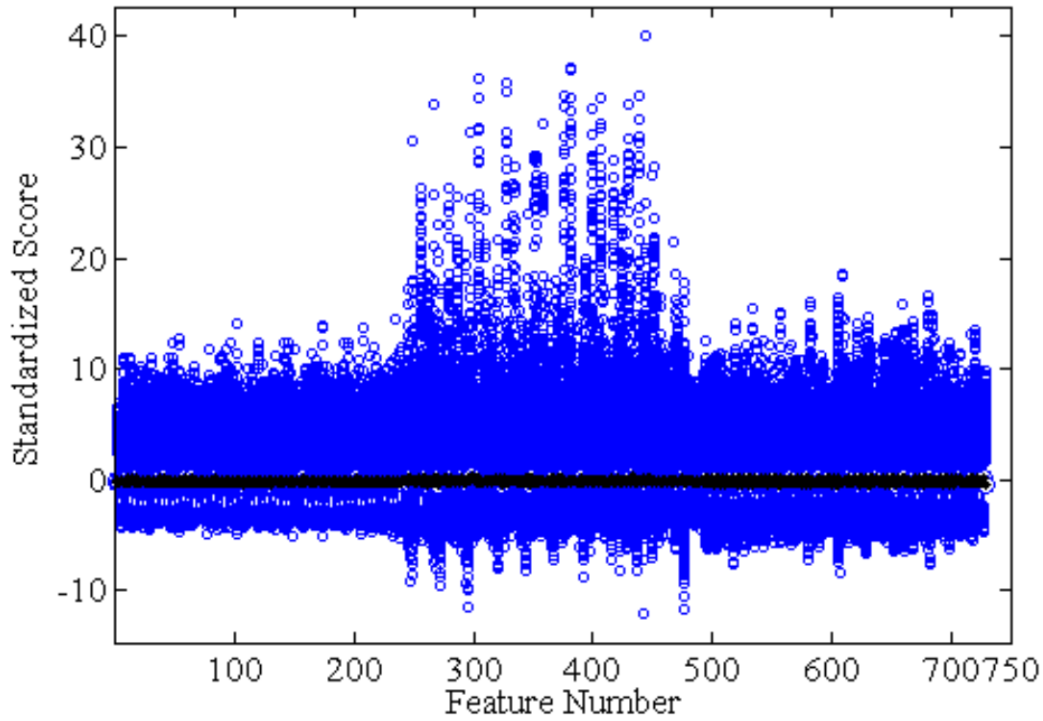


Figure VI-2: Boxplots of ZigBee RF-DNA Features at 10dB for Authorized Devices Using Standardized Data.

When using standardized RF-DNA features with MDA/ML processing, negligible gains (G), the reduction in required SNR expressed in dB to achieve a given $\%C$, of $G = 0.09$ dB (TNG) and $G = 0.06$ dB (TST) are realized at $\%C = 90\%$ when compared with nominal centered and maximum scaled RF-DNA features. Thus there is effectively no difference between performance outputs and the slight differences are logically assignable to differing random values used in the Additive White Gaussian Noise (AWGN) process.

While this shows a negligible impact of standardization on classification performance, the normalization and standardization method may still be important. It is

possible that a DRA and standardization could lend itself to improved performance. Therefore, DRA using a low number of features was pursued; Unscaled Summed MDA Loadings Fusion (*USum* MLF), Section 4.2.3.1, was used to select 10 features. With the top 10 features, classification accuracy does not achieve $\%C = 90\%$, however one can potentially get determine features very useful for discrimination, as discussed in Section IV. With $N_{DRA} = 10$, classification accuracy was evaluated using MDA/ML models, with Relative Accuracy Percentage (*RAP*) values are computed with respect to the nominal TST MDA/ML model; between the nominal and standardized approaches of $RAP_{TNG} = 0.970$ (TNG) and $RAP_{TST} = 0.968$ (TST) were computed, indicating that the nominal approach offers higher accuracy. Therefore, empirically, the nominal mean centered and max-scaled RF-DNA normalization has a small, but distinct, advantage over standardization.

6.3 Simulation Methods, Dependence and Correlation Effects in RF-DNA

Simulation is a tool used by operations research professionals to model and understand complex processes [136]. One area of interest in simulation research is examining steady state conditions from a time sampled output. One commonly then divides a steady state signal into independent and uncorrelated batches. The batches are then examined to provide insight into how a given system functions. Particular emphasis will be given towards signal autocorrelation to determine batch sizes, data standardization, and batching means methods that leverage knowledge of the signal itself and the binning process. Simulation studies involve collecting input and output data, and

parameters to create a statistical model of a real or hypothetical system under analysis [136]. Simulation research is prominent in engineering, business and operations research applications (cf. [136, 243, 544–548]). Simulations can involve multiple short sets of system output data or one long-run of system output data. When one long set of observation data is available and it is prohibitive to collect additional data, batches are one approach used to provide additional data about steady state condition [136]. Batches are constructed as visualized in Figure VI-3 as M subregions of the sample, with each region considered as a separate observation and containing N_S samples per subregion [243]. The 0^{th} batch in Figure VI-3 is considered a transient region and is not used for analysis. To ensure that each batch can be considered as a separate observation of the system in steady state, understanding the independence and correlation of batches is needed [136]. Additionally, when analyzing simulation data, one first needs to identify the point where the system reaches steady state and is not influenced by startup characteristics [136]. In other business analytics domains, similar approaches to batching are termed binning, c.f. [199, 549].

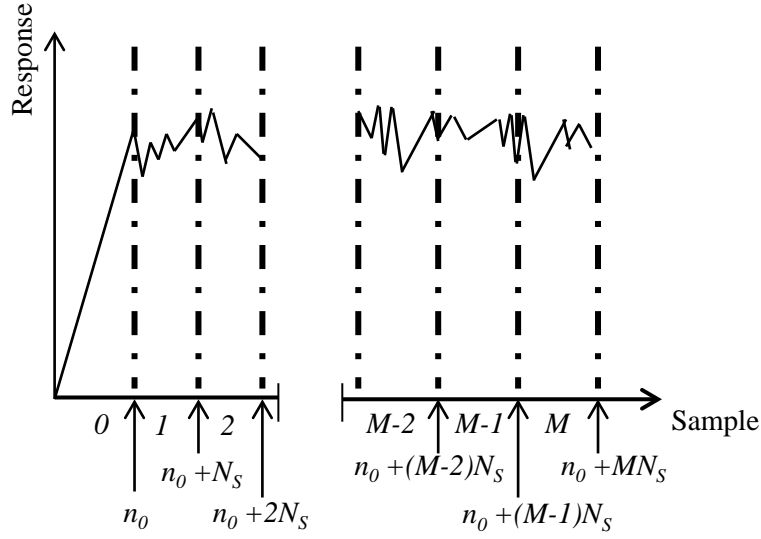


Figure VI-3: General Batching Method for Simulation Output Showing the Response Divided into M Total Batches [243].

The batching process in simulation parallels closely with the RF-DNA Fingerprinting process, as described in Section 2.4, in that a signal's Regions of Interest (ROI) is divided into N_R equally sized subregions which are then processed for further analysis. Since the RF-DNA process yields distinction between devices, it is logical that RF-DNA fingerprints computed from independent measurements will be useful for device discrimination. Therefore, methods from simulation aimed at reducing correlation effects in the data could be beneficial to RF-DNA.

6.3.1 Transient Determination

Transient periods are present as a system begins to operate, in Figure VI-3 the transient period is batch 0. Transients (considered as startup biases) are detrimental to simulation studies [550], thus simulation studies generally desire to consider only steady-

state processes in order to accurately model a process by reducing influence of startup characteristics [136]. Automated transient detection methods were proposed by [551]. Objectively, transient determination is similar to discarded initialization region in RF-DNA. While automated approaches for transient determination could be applied in the RF-DNA Fingerprinting process, such approaches are not considered herein since the ROI is device dependent. However, future work may wish to examine this area in conjunction with leveraging knowledge about the communication signal itself to determine and isolate the ROI for RF-DNA.

6.3.2 Autocorrelation and the Number of Batches

Batch size is another important question in simulation analysis [552–554]. Additionally, higher order moments (such as 3rd and 4th order) were determined by [555] to be more sensitive to interval differences than lower order moments. Therefore, selecting appropriately sized ROIs may be critical to RF-DNA device classification and device ID verification performance. In simulation studies, normality of a given batch can be one factor used to determine batch size [556]. Various approaches (in multiple disciplines) exist, e.g. [199, 549, 557–562], for determining batch size. Determining the appropriate number of batches to create minimizes correlation between batches, see [243, 551, 554], and it is of interest to produce independent batches.

Although inter-feature correlation can be beneficial to classification performance, intra-feature correlation (correlation between data features) generally causes adverse effects to classification performance [563–565]. The reasoning for this is that highly

correlated features are redundant [566–568]. In other words, if $\text{corr}(\mathbf{X}, \mathbf{Y}) = 1$, then $\text{corr}(\mathbf{X}, \mathbf{Y}) = \text{corr}(\mathbf{X}, a\mathbf{X}) = \text{corr}(\mathbf{X}, \mathbf{X})$, indicating that no information was added by retaining both features. Multiple correlated features can also cause instability issues in linear methods such as ANOVA, logistic regression, linear least squares regression, and discriminant analysis [564, 568]. While nonlinear classifiers can process correlated data, e.g. [569], redundant features will still increase computation time and are undesirable [567].

The covariance between two variables \mathbf{X} and \mathbf{Y} is defined as

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = \text{Cov}(\mathbf{Y}, \mathbf{X}) = E[(\mathbf{X} - E(\mathbf{X}))E[(\mathbf{Y} - E(\mathbf{Y}))]], \quad (6.2)$$

with the correlation of \mathbf{X} and \mathbf{Y} being the scaled covariance,

$$\text{Corr}(\mathbf{X}, \mathbf{Y}) = \frac{\text{Cov}(\mathbf{X}, \mathbf{Y})}{\sqrt{\text{Var}(\mathbf{X})}\sqrt{\text{Var}(\mathbf{Y})}} \quad (6.3)$$

which normalizes the covariance to have values between -1 and $+1$ [551].

To consider batch means and autocorrelation computations we considering a generic steady-state sequence vector \mathbf{V}_n for $n = 1, 2, \dots, N$ [243, 551], where N is the total number of samples. For this sequence vector, we compute the steady-state mean as

$$E[\mathbf{V}_n] = \frac{\sum_{i=1}^N \mathbf{V}_i}{N} = \boldsymbol{\mu} , \quad (6.4)$$

and variance as

$$E[(\mathbf{V}_n - \boldsymbol{\mu})^2] = \boldsymbol{\sigma}^2 . \quad (6.5)$$

The autocorrelation function for a sequence vector is a covariance function with properties,

$$\gamma(0) = \sigma^2 \quad (6.6)$$

$$\gamma(K) = \gamma(-K)$$

where K is an offset [551]. Of interest is determining the spacing within a sequence to find the covariance stationary quantity

$$\text{Cov}(\mathbf{V}_n, \mathbf{V}_{n+K}) = \gamma(K), \quad (6.7)$$

for any n and K [551]. With these quantities, dependence can be computed via the correlation, where (6.3) is computed for,

$$\rho(K) = \text{Corr}(\mathbf{V}_n, \mathbf{V}_{n+K}) = \frac{\text{Cov}(\mathbf{V}_n, \mathbf{V}_{n+K})}{\sqrt{\text{Var}(\mathbf{V}_n)}\sqrt{\text{Var}(\mathbf{V}_{n+K})}} = \frac{\gamma(K)}{\gamma(0)} \quad (6.8)$$

which is the correlation within the sequence with a separation of K [551]. Correlation has various interesting and useful properties,

$$\rho(0) = 1$$

$$\rho(K) = \rho(-K) \quad (6.9)$$

$$-1 \leq \rho(K) \leq 1.$$

In RF-DNA fingerprinting, one extends this process by realizing that K is equivalent to N_S , the total number of samples in a subregion. Since N_R , the total number of subregions, are frequently empirically determined, N_S is also empirically determined in prior RF-DNA work, see [18, 59, 89]. However, autocorrelation could assist in this process by determining the number of time samples-per-subregion which lead to uncorrelated subregions. When computing the autocorrelation function for multiple devices, one aims to find the number of time samples-per-subregion associated with the smallest autocorrelation. For multiple devices, one should simultaneously compare the

autocorrelation function of all devices with the best minimum autocorrelation across device used to determine ROI size.

The autocorrelation amplitude for the 4 authorized ZigBee devices is presented in Figure VI-4 at $SNR = 10$ dB, along with a line of 0 autocorrelation. Of interest in Figure VI-4 is when the autocorrelation functions are at the 0 autocorrelation line, which indicates minimum autocorrelation. Figure VI-4 shows that minimum autocorrelation (approximately 0) for the four devices occurs at autocorrelation indices of 24 time samples-per-subregion and 48 time samples-per-ROI. Incidentally 48 time samples-per-subregion and 24 samples-per-subregion correspond, respectively, with 1 subregion-per-bit and 2 subregions-per-bit as explored by Dubendorfer [91]. While Dubendorfer [91] employed a physical understanding of signal structure and findings of prior empirical work to determine subregion size, employing autocorrelation for ROI size determination adds robustness to this decision.

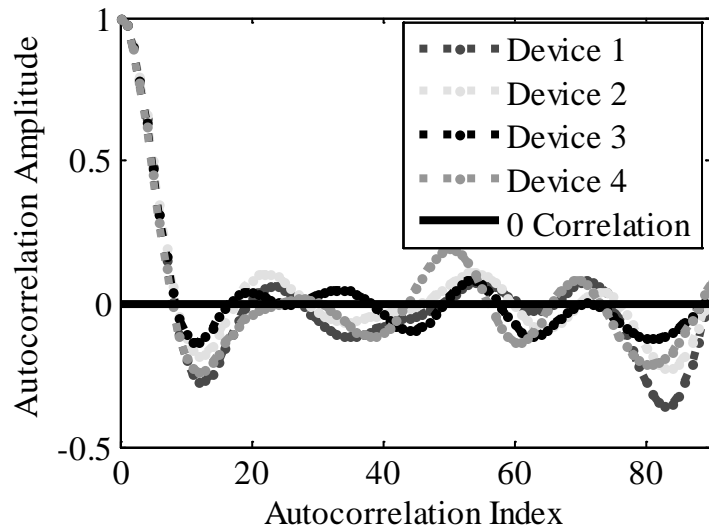


Figure VI-4: Autocorrelation of ZigBee Data Features, $SNR = 10$ dB.

VII. Summary and Conclusions

What means all this?

–MARCUS AURELIUS, 121-180

This document presents various theoretical, practical, and application-based contributions made in the Radio Frequency (RF) Fingerprinting arena, including advancements in classifier model development, Dimensional Reduction Analysis (DRA), and AFIT's RF Distinct Native Attribute (RF-DNA) Fingerprinting process. This chapter presents a summary of the research, its contributions and recommendations for future research.

7.1 Research Summary

Simple, low-cost wireless devices permeate the world, including those used in Critical Infrastructure (CI) applications where they interact with physical devices. ZigBee and Z-Wave devices are two devices and have well-known security issues --c.f. [37, 38, 170] and are of interest for this research. When considering security and a hierarchy of communication signaling, such as the seven layer Open System Interconnection (OSI) model [62–64], security is generally only considered within the Application, Network and Data Link layers [51–58]. Much less emphasis has been placed on Physical (PHY) layer security, the interface layer of signals emanating from the device itself, and extensions of PHY-based RF-DNA Fingerprinting process are of interest for improving security.

RF-DNA Fingerprinting aims to exploit device emissions in a biometric-like manner where statistical features having attributes of *universality*, *distinctiveness*, *permanence*, and *collectability* are generated and used for *Device Classification* and *Device ID Verification* [19, 66]. RF-DNA fingerprints are statistical in nature and involve computing the variance, skewness and kurtosis within Regions of Interest (ROI) selected from instantaneous amplitude, phase, and frequency responses. When considering RF-DNA fingerprints, one must develop a classifier model to discriminate between devices. Previous efforts have introduced and employed Multiple Discriminant Analysis (MDA), Generalized Relevance Learning Vector Quantization Improved (GRLVQI), Random Forests, and Learning From Signals (LFS) [51, 90, 133, 134] processes for classification. Herein, the MDA and GRLVQI processes are considered and extended. Additionally, RF-DNA features are frequently numerous and thus DRA is of interest to select appropriate subsets of features. Prior DRA research in RF-DNA has considered the two-sample Kolmogorov-Smirnov (KS) test and GRLVQI relevance ranking values. Herein, multiple extensions to DRA were made to introduce new methods, develop an MDA-based DRA method, and improve the understanding of DRA methods.

Deficiencies in p -value based DRA were illustrated and the proposed F-test and revised KS-test illustrated advantages in using test statistic values for DRA. Further improvements in DRA included developing quantitative dimensionality assessment DRA was shown to remove subjectivity when selecting DRA subsets. MDA-based Loadings Fusion (MLF) was shown to be an MDA-classifier based DRA method which resolved

previously mentioned deficiencies in MDA [51, 91, 92, 113, 134, 241]. The proposed autocorrelation-based approach to RF-DNA fingerprint subregion size specification was shown to add robustness to the previously subjective RF-DNA fingerprinting subregion specifications.

The proposed F-test and MLF DRA methods were shown to offer distinct performance improvements over the KS-test and GRLVQI DRA methods. ZigBee *Device Classification* results for selected DRA methods with an MDA/ML classifier and arbitrary average correct classification (%C) benchmark of %C = 90%, included SNR gain (G_{SNR}) relative to the benchmark GRLVQI DRA with $N_{DRA} = 50$ feature sets of 1) $G_{SNR} = +0.82$ dB for *SSum* MLF DRA, and 2) $G_{SNR} = +0.10$ dB for F-test DRA using $N_{DRA} = 50$, compared to 3) $G_{SNR} = +0.71$ dB for KS-Test DRA using $N_{DRA} = 50$, and 4) $G_{SNR} = -4.22$ dB for the baseline Random DRA using $N_{DRA} = 50$. ZigBee *Device ID Verification* results, using the same $N_{DRA} = 50$ feature sets and MDA/ML classifier, included correct verification of *authorized* device IDs (% V_A) and correct detection of unauthorized *rogue* device IDs (% V_R) of % $V_A = 50\%$ % $V_R = 91.67\%$ for the benchmark GRLVQI DRA, with 1) % $V_A = 50\%$ and % $V_R = 91.67\%$ for *SSum* MLF DRA, and 2) % $V_A = 75\%$ and % $V_R = 91.67\%$ for F-test DRA, compared to 3) % $V_A = 50\%$ and % $V_R = 86.11\%$ for the KS-test, and 4) % $V_A = 50\%$ and % $V_R = 75\%$ for the baseline Random DRA. Thus the proposed *SSum* MLF DRA and F-Test DRA offer a performance advantage over both GRLVQI DRA and KS-Test DRA while being computationally and conceptually simpler DRA methods.

The optimized GRLVQI algorithm and the proposed GRLVQI-D algorithm showed improved performance over the baseline GRLVQI algorithm. When considering GRLVQI classifier improvements using $N_F = 189$ Z-Wave features and the $\%C = 90\%$ benchmark, demonstrated *Device Classification* performance relative to baseline GRLVQI using a squared-Euclidean distance measure includes 1) improved $G_{SNR} = +1.94$ dB using the GRLVQI optimized algorithm, and 2) improved $G_{SNR} = +1.84$ dB using GRLVQI-D with a Cosine distance measure. For Z-Wave *Device ID Verification*, results include 1) worst case $\%V_A = 33.33\%$ for baseline GRLVQI, 2) improved $\%V_A = 66.66\%$ for GRLVQI-D using a Cosine distance measure, and 3) best case $\%V_A = 100\%$ using the optimized GRLVQI algorithm. Due to availability, Z-Wave devices were not present for *rogue* device assessments. When ZigBee RF-DNA fingerprints were considered using the Z-Wave optimized GRLVQI and GRLVQI-D algorithms, performance was worse than the nominal settings of Reising [51], indicating that the Z-Wave optimal settings are not applicable to ZigBee device discrimination.

7.2 Research Contributions

Three primary contributions were made under this research, including improvements to 1) the Dimensional Reduction Analysis (DRA) methodology, 2) the GRLVQI classifier, and 3) the RF-DNA Fingerprinting process. A summary of each follows:

1. **DRA Improvements:** Includes development and analysis of MDA Loadings Fusion (MLF) methods to rectifying the reported issue in, c.f.

[51, 91, 92, 113, 134, 241], that includes MDA lacking a classifier-based relevance. An F-test DRA method was introduced and shown to offer reasonable performance. Quantitative DRA assessment methods were developed to determine the number of retained features (N_{DRA}) and their performance compared with previous qualitative DRA methods of [91]. Prior RF-DNA DRA efforts have considered p -values for feature relevance ranking [89, 113]. However, phenomenological issues exist with such an approach, an improved understanding is developed herein based on the merits of p -values versus test statistics for feature relevance ranking. Finally, a preliminary investigation into DRA relevance fusion was presented.

2. GRLVQI Classifier Improvements: involved changing the underlying distance measure. To do so, one must necessarily change the cost function and derivatives to the GRLVQI algorithm. Since a) GRLVQI is a rather complicated algorithm and b) many different distance measures exist, a procedure to select different distance measures was created that involved first comparing distance measures themselves and then iteratively incorporating a distance measure into successively more complicated learning vector quantization (LVQ) algorithms leading up to GRLVQI. For this process, the first known derivative framework for the LVQ-family of algorithms was developed. Subsequently, an optimization approach was presented to determine reasonable algorithm parameter

settings for the baseline GRLVQI process and the newly developed distance-based GRLVQI process (GRLVQI-D).

3. RF-DNA Fingerprinting Improvements: An enhanced understanding of the nature of instantaneous amplitude, phase, and frequency features was developed to better understand why phase features have historically been the most relevant for device classification. An autocorrelation method was developed and characterized to automate the determination of the number of subregions used within a given response ROI. Finally, a first-look assessment of simulation-based ROI weighting schemes was completed for RF-DNA Fingerprinting.

7.3 Proposed Future Research

Given the methods developed under this research and corresponding findings, many different future research endeavors could be pursued. The following are proposed:

1. **Additional GRLVQI Algorithm Extension:** Herein, distance measures and the relative distance difference equation were changed in the GRLVQI algorithm. However, future work could consider different activation functions, e.g. [570], to replace the sigmoid operation in GRLVQI. The presented LVQ-family derivative skeleton would be an initial starting point in this effort.
2. **Tailor Algorithmic Optimization to the Signal of Interest:** Optimizing the GRLVQI algorithmic settings was considered for Z-Wave data and

shown to be viable. When these settings were applied to the ZigBee dataset, performance was degraded relative to the baseline. To compute optimal settings for the ZigBee dataset, one would require many algorithmic runs which would be computationally costly. To facilitate large-scale algorithm optimization studies, employing the Air Force Research Laboratory DOD Supercomputing Resource Center (DSRC) should be considered. Employing DSRC would facilitate tailored GRLVQI settings to a given signal of interest, in addition to permitting comparing different optimization methods.

3. **Extend DRA Methods:** Herein, two additional DRA methods (F-test and MDA Loadings) were introduced for RF-DNA Fingerprinting applications. Additional DRA methods are identified in literature and could be considered, including entropy [76], Best Individual Features [213, 571], Logistic Principal Component Analysis (PCA) [572], nonlinear PCA [213], kernel PCA [213], and Independent Component Analysis (ICA) [213, 573].
4. **Revisit DRA Fusion:** The DRA fusion methods considered herein demonstrated some utility at lower N_{DRA} values. This could be explored further to identify other alternate DRA fusion schemes.
5. **Further Consider Simulation Methods:** Autocorrelation methods from Simulation were shown to be applicable to RF-DNA Fingerprinting. Additional Simulation methods that consider weighting distributions to

reduce correlation effects, e.g. [136, 544–546, 574–578], could be developed and applied to Region of Interest (ROI) subregions.

6. **Explore RF-DNA Feature Phenomenology:** It was seen that instantaneous phase features are generally more relevant than both amplitude or frequency features and some insight was developed to address this. However, to better understand the relationship between feature type and their relevance to the classification decision, additional studies could be performed. In this case, one could consider simulated devices (agnostic of modulation) and similar devices that differ only by the modulation they employ.

7.4 Sponsor Acknowledgement

This research was supported in part by the U.S. Air Force Research Laboratory, Sensors Directorate (AFRL/RY), Wright-Patterson Air Force Base, OH. Additional support was provided by the Air Force Institute of Technology (AFIT), including the Center for Cyberspace Research (CCR) and Center for Technical Intelligence Studies and Research (CTISR).

APPENDIX A: Lemma Associated with Multiple Discriminant Analysis Loadings

Learning is essentially hard; it happens best when one is deeply engaged in hard and challenging activities.

–SEYMOUR PAPERT, 1928 –

Lemma 1: if a is a scalar, b is a vector, and X is a matrix, then if one is computing the correlation of $ab^T X$ and X then $\text{corr}(X, ab^T X) = \text{corr}(X, b^T X)$.

To prove Theorem 1, the scaling will be represented as eigenvectors

$$b^* = ab, \quad (\text{A.1})$$

scaled by a scalar a [237]. If the projection matrix were scaled, as in (A.1), then the relationship in (3.11) could be expressed as

$$\text{corr}(X, b^{*T} X) = \text{corr}(X, X) D_X^{1/2} b^* [b^{*T} \text{cov}(X, X) b^*]^{-1/2}, \quad (\text{A.2})$$

which expands to

$$\text{corr}(X, ab^T X) = \text{corr}(X, X) D_X^{1/2} ab [ab^T \text{cov}(X, X) ab]^{-1/2}. \quad (\text{A.3})$$

Equation (A.3) can be expanded to

$$\begin{aligned} \text{corr}(X, X) D_X^{1/2} aba^{-1} [(b^T \text{cov}(X, X) b)]^{-1/2} \\ = \text{corr}(X, ab^T X), \end{aligned} \quad (\text{A.4})$$

which means the scaling multiplier can cancel, yielding the conclusion that scaling the loadings does not change the loadings,

$$\text{corr}(X, ab^T X) = \text{corr}(X, b^T X). \quad (\text{A.5})$$

APPENDIX B: Examination of LVQ and GLVQ Properties and Features

You can't process me with a normal brain...

–CARLOS ESTÉVEZ, 1965 –

The GLVQ, GRLVQ, and GRLVQI relative distance measure in (3.34) deserves some understanding of what this actually measures. A simple example can be constructed with a simple example. Consider a hypothetical space presented in Figure B-1 where there are two hypothetical PVs, placed at $(-1, 1)$ and $(-1, -1)$ respectively, and an exemplar at $(1, 1)$. The squared Euclidean distances between the exemplar and each PV are respectively

$$d_{PV_1} = 4 \quad (\text{B.1})$$

and

$$d_{PV_2} = 8. \quad (\text{B.2})$$

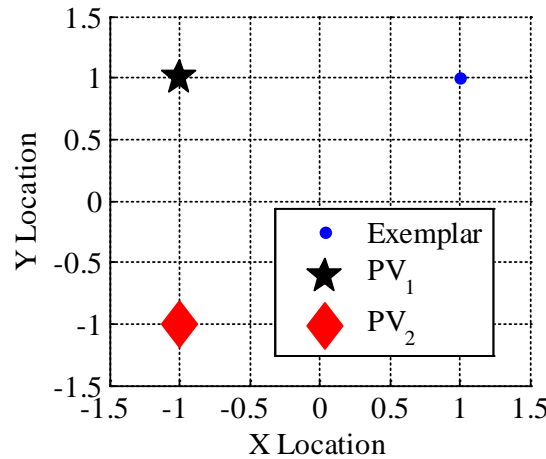


Figure B-1: Hypothetical Situation with Two PVs and One Exemplar

To consider the output of the relative distance measure in (3.34), one can consider two situations, 1) PV_1 being the correct in-class PV or 2) PV_2 being the correct in-class PV. For case 1), the relative distance difference measure returns a score of -0.3333 , but in 2) the relative distance difference measure returns a score of 0.3333 . Per the discussion in Section 3.3.1.6 on interpreting the distance difference measure, negative values are indicative of correct classification and positive values are incorrect classification with the magnitude indicating how “correct” or “incorrect.”

To extend this example of how the PVs, exemplar, distance measure, and relative distance difference interact, one can extend this example to compute the distance of every point to the two stationary PVs. Figure B-2 presents the squared Euclidean distance for every point (0.01 sampling) between -4 and 4 and the two PVs. Figure B-2a presents the values where PV_1 is considered, and Figure B-2b presents the values where PV_2 is considered. Logically, the distances form circles of increasing distance from the respective PVs.

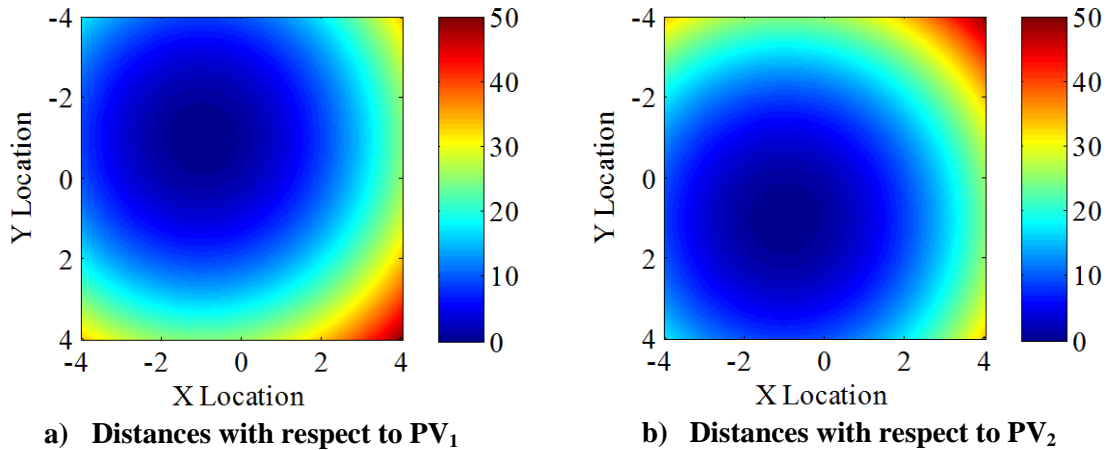


Figure B-2: Distances Between Exemplars and a) PV_1 and b) PV_2

Considering the relative distance difference metric for Figure B-2, and assuming the PV_1 is the correct classification, one sees Figure B-3. Here one can see that the scores go to -1 as one approaches PV_1 and $+1$ as one approaches PV_2 with curves of different values around each PV. As PV_1 and PV_2 move closer together, one finds that most possible points for an exemplar are scored near 0, while only scores extremely close to each PV receive higher magnitude scores, as seen in Figure B-4.

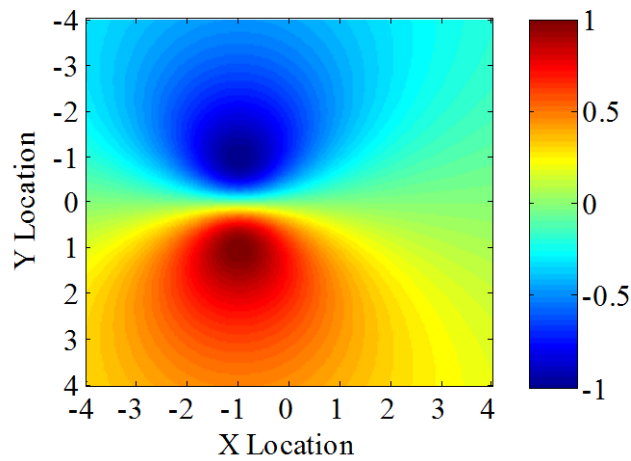


Figure B-3: General Relationship Between Distance Difference Measure and PV Distances

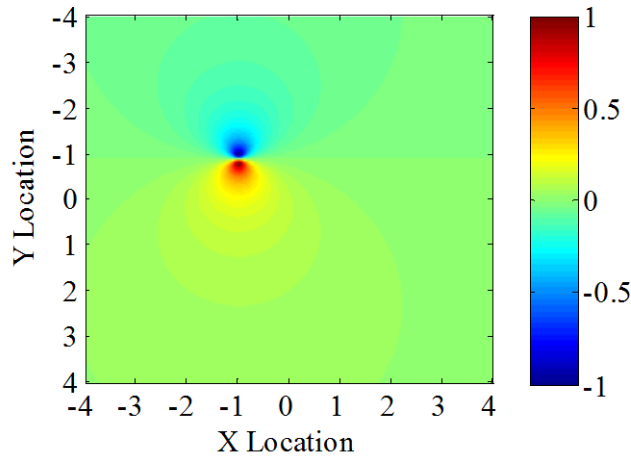


Figure B-4: Relationship Between Distance Difference Measure and PV Distances for Closely Spaced PV_1 and PV_2

APPENDIX C: *P*-values versus Test Statistics on Selected Academic Datasets

*...the primary product of a research inquiry is one or more measures of effect size, not *p* values.*

–JACOB COHEN, 1923 – 1998

Section 4.2.1.3 showed that *p*-values were largely deficient as a feature relevance ranking tool for RF-DNA due to *p*-values 1) being computed beyond machine precision, 2) having less resolution than test statistic values, 3) converging on zero, and 4) offering slightly less classification performance than test statistic relevance ranking. However, this was only a single example on a specific problem; therefore this appendix presents empirical demonstrations on academic datasets to show that this problem is not unique to RF-DNA.

To examine the generalizability of *p*-value versus test statistic feature ranking, a selection of academic datasets was examined as presented in Table C-1. Table C-1 presents a consistent amount and variety of data as examined in [579]. The datasets consist of well-known multivariate problems and range in size from 30 exemplars, 3 features, and 3 classes in Insect to 60,000 exemplars, 717 features, and 10 classes in MNIST.

All datasets were considered using the KS-test and F-test feature relevance ranking methods, consistent with Section 4.2.1.3. To compute *p*-values, with the exception of MNIST, no separation into training and testing sets were pursued and all datasets were considered in their entirety.

Table C-1: Example Academic Datasets.

DATASET	NUMBER OF SAMPLES IN CLASSES		NUMBER OF FEATURES	TOTAL NUMBER OF EXEMPLARS	REFERENCE
FISHER	Setosa: 50 Versicolor: 50 Virginica: 50		4	150	[235, 580]
INSECT	Species 1: 10 Species 2: 10 Species 3: 10		3	30	[466, 467]
VERTEBRAL COLUMN	Spondylolisthesis: 150 Normal: 100 Disk Hernia: 60		6	310	[581]
WINE QUALITY	White: 4898 Red: 1599		11	6497	[582]
WISCONSIN BREAST CANCER	Benign: 458 Malignant: 241		9	699	[583]
WINE	Cultivar 1: 59 Cultivar 2: 71 Cultivar 3: 48		13	178	[584]
MNIST (TRAINING SET)	1: 6742 2: 5958 3: 6131 4: 5842 5: 5421	6: 5918 7: 6265 8: 5851 9: 5949 0: 5923	784	60,000	[585, 586]
ECOLI	Cytoplasm: 143 Inner Membrane: 116 Periplasm: 52 Outer Membrane: 25		7	336	[587]

Fisher Iris was first examined using the p -value and test statistic approaches described in Section 4.2.1.3. The Fisher Iris dataset is a commonly used academic

discrimination problem that contains measurements of petals and sepals for three species of Iris flowers: setosa, versicolor, and virginica. This dataset contains 50 observations per class, no missing values, and four data features: petal length, petal width, sepal length, and sepal width [235]. Table C-2 presents a similar comparison of features as in Table IV-2; however, since Fisher Iris consists of only 4 features the features are not sorted and the test statistic values represent the actual values for those features. Again, as in Section 4.2.1.3, many p -values were computed as values beyond machine zero while their associated test statistic values are reasonable.

Table C-2: p-values vs Test Statistic for Fisher Iris

FEATURE NUMBER	F-TEST		KS-TEST	
	TEST STATISTIC	P-VALUE	SUMMED TEST STATISTIC	SUMMED P-VALUE
1	119.26	$1.67 \cdot 10^{-31}$	9.400	$1.74 \cdot 10^{-21}$
2	49.16	$4.49 \cdot 10^{-17}$	2.4733	$1.68 \cdot 10^{-22}$
3	1,180.20	$2.86 \cdot 10^{-91}$	1.800	$1.91 \cdot 10^{-21}$
4	960.00	$4.17 \cdot 10^{-85}$	2.5733	$2.84 \cdot 10^{-30}$
VARIANCE	332,880.0	$5.04 \cdot 10^{-34}$	12.7836	$1.02 \cdot 10^{-42}$

The Insect data considers three species, 10 observations each with no missing values, of *chaetocnema* insects [499, 500]. Data feature here correspond to: width of the first joint of the first tarsus (microns), width of the first joint of the second tarsus (microns), and maximal width of the aedeagus (microns) [499, 500]. While no p -values below machine precision were computed, Table C-3 shows again the value of test-

statistic ranking over p -value ranking since the differences between KS-test p -values are very small to be imperceptible.

Table C-3: p -values vs Test Statistic for Insect.

FEATURE NUMBER	F-TEST		KS-TEST	
	TEST STATISTIC	P-VALUE	SUMMED TEST STATISTIC	SUMMED P- VALUE
1	64.88	0.00	2.77	$1.11 \cdot 10^{-8}$
2	1.36	0.27	1.77	$1.11 \cdot 10^{-8}$
3	1.12	0.34	2.0	$3.59 \cdot 10^{-14}$
VARIANCE	1,350.1	0.033	0.27	$4.09 \cdot 10^{-17}$

The vertebral column dataset considers spine measurements and normal and abnormal disk issues, such as Disk Hernia and Spondylolisthesis [584]. When examining the vertebral column dataset, Table C-4, many p -values are seen as being computed beyond machine precision. However, the test statistic values offer more perceptible differences between features.

Wine Quality considers various chemical properties, e.g. acidity and sulphates, in the Portuguese "Vinho Verde" wine and their relationship with a quality score [582]. Table C-5 presents results for the KS-test and F-test DRA approaches; while all but two KS-test summed p -values were equal to exactly zero with the non-zero values being below machine precision, the KS-test statistic value offers a seemingly reasonable approach to rank features. A similar result is also seen in the F-test for this data.

Table C-4: p-values vs Test Statistic for Vertebral Column.

FEATURE NUMBER	F-TEST		KS-TEST	
	TEST STATISTIC	P-VALUE	SUMMED TEST STATISTIC	SUMMED P- VALUE
1	98.537	$8.77 \cdot 10^{-34}$	3.129	$4.88 \cdot 10^{-7}$
2	21.298	$2.22 \cdot 10^{-9}$	3.787	$3.42 \cdot 10^{-16}$
3	114.988	$5.34 \cdot 10^{-38}$	2.777	$4.89 \cdot 10^{-7}$
4	89.647	$2.17 \cdot 10^{-31}$	2.923	$1.41 \cdot 10^{-9}$
5	16.869	$1.12 \cdot 10^{-7}$	4.823	$4.69 \cdot 10^{-122}$
6	119.127	$5.10 \cdot 10^{-39}$	3.013	$3.42 \cdot 10^{-16}$
VARIANCE	2,111.9	$2.07 \cdot 10^{-15}$	0.602	$6.36 \cdot 10^{-14}$

Table C-5: p-values vs Test Statistic for Wine Quality.

FEATURE NUMBER	F-TEST		KS-TEST	
	TEST STATISTIC	P-VALUE	SUMMED TEST STATISTIC	SUMMED P- VALUE
1	8.00	$1.26 \cdot 10^{-8}$	9.31	0.0
2	96.67	$8.44 \cdot 10^{-117}$	8.64	$9.86 \cdot 10^{-21}$
3	9.31	$3.44 \cdot 10^{-10}$	8.58	$9.86 \cdot 10^{-21}$
4	9.11	$5.97 \cdot 10^{-10}$	8.48	0.0
5	50.85	$1.95 \cdot 10^{-61}$	9.84	0.0
6	14.94	$4.77 \cdot 10^{-17}$	9.17	0.0
7	7.72	$2.77 \cdot 10^{-8}$	9.66	0.0
8	136.95	$6.58 \cdot 10^{-164}$	9.96	0.0
9	2.02	0.06	9.48	0.0
10	4.33	$2.31 \cdot 10^{-4}$	9.19	0.0
11	320.59	0.0	9.45	0.0
VARIANCE	9,434.7	$3.19 \cdot 10^{-4}$	0.25	$1.59 \cdot 10^{-6}$

The Wisconsin Breast Cancer dataset concerns various parameters about potential breast masses for a classification of benign or malignant [583]. As seen in the other examples, Table C-6 presents results for the KS-test and F-test DRA approaches. Again, for both approaches, test statistic values are seen to provide results which are real numbers and not beyond machine precious or infinitesimally small.

Table C-6: p-values vs Test Statistic for Wisconsin Breast Cancer.

FEATURE NUMBER	F-TEST		KS-TEST	
	TEST STATISTIC	P-VALUE	SUMMED TEST STATISTIC	SUMMED P- VALUE
1	733.21	$6.84 \cdot 10^{-111}$	3.05	$7.92 \cdot 10^{-21}$
2	1,408.5	$1.75 \cdot 10^{-169}$	1.744	0.66
3	1,419.3	$2.95 \cdot 10^{-170}$	1.743	0.51
4	657.79	$1.11 \cdot 10^{-102}$	1.84	0.48
5	608.72	$4.35 \cdot 10^{-97}$	3.78	$9.40 \cdot 10^{-9}$
6	1,014.2	$4.54 \cdot 10^{-138}$	2.02	$1.18 \cdot 10^{-4}$
7	933.29	$9.85 \cdot 10^{-131}$	2.79	$9.40 \cdot 10^{-9}$
8	717.63	$3.12 \cdot 10^{-109}$	1.99	0.32
9	152.04	$9.68 \cdot 10^{-32}$	3.30	$4.51 \cdot 10^{-12}$
VARIANCE	160,430	$1.04 \cdot 10^{-63}$	0.59	0.07

The wine dataset is conceptually similar to the wine quality dataset, however here we are interested in discriminating between three different grape cultivars [584]. Similar to the other example datasets, p -values are again computed beyond machine precision and offer less obvious interpretability as that seen in the test statistic values. However, one issue does exist in the KS-test statistic values with feature 5 and 13 producing identically valued test statistics, but this is the only occurrence of this problem and

despite this issue the test statistic values still appear to offer more consistent and interpretable relevance ranking values.

Table C-7: p-values vs Test Statistic for Wine

FEATURE NUMBER	F-TEST		KS-TEST	
	TEST STATISTIC	P-VALUE	SUMMED TEST STATISTIC	SUMMED P- VALUE
1	135.07	$3.32 \cdot 10^{-36}$	11.93	$3.18 \cdot 10^{-71}$
2	36.94	$4.13 \cdot 10^{-14}$	7.94	0.002
3	13.31	$4.15 \cdot 10^{-6}$	9.19	$5.978 \cdot 10^{-7}$
4	35.77	$9.44 \cdot 10^{-14}$	11.92	$3.18 \cdot 10^{-71}$
5	12.43	$8.96 \cdot 10^{-6}$	12.00	$7.99 \cdot 10^{-79}$
6	93.73	$2.14 \cdot 10^{-28}$	8.12	$3.11 \cdot 10^{-4}$
7	233.93	$3.59 \cdot 10^{-50}$	7.73	0.0017
8	27.58	$3.88 \cdot 10^{-11}$	11.75	$2.96 \cdot 10^{-62}$
9	30.27	$5.13 \cdot 10^{-12}$	8.92	$1.79 \cdot 10^{-7}$
10	120.66	$1.16 \cdot 10^{-33}$	10.19	$1.48 \cdot 10^{-27}$
11	101.32	$5.92 \cdot 10^{-30}$	10.95	$1.20 \cdot 10^{-35}$
12	189.97	$1.39 \cdot 10^{-44}$	8.52	$4.45 \cdot 10^{-6}$
13	207.92	$5.78 \cdot 10^{-47}$	12.00	$7.99 \cdot 10^{-79}$
VARIANCE	6,040.10	$7.03 \cdot 10^{-12}$	3.02	$4.77 \cdot 10^{-7}$

Written character recognition is a concern in many fields, e.g. [588–592], MNIST is a dataset that considers thousands of handwritten digits [585, 586]. MNIST’s data features are actually pixels in an 28x28 image, with each of the 60,000 observations containing one image of one handwritten digit [585, 586]. However, the final image is really 20x20 since there is a band of 0s around the 20x20 image [585, 586]. Table C-8 presents results when the KS-test and F-test DRA approaches are applied. Values are

sorted from lowest to highest based on the respective test-statistic value, consistent with those presented for RF-DNA. Notably, this dataset shows that the F-test failed to produce a test statistic value in some cases, while the KS-test did not. However, underlying this issue is the data itself; many observations in some features were all 0s, therefore such a result is understandable since the KS-test is comparing two distributions and the distributions of two vectors of all zeros is identical. Therefore, the KS-test has no issue with handling such data, while the F-test does.

Table C-8: p-values vs Test Statistic for MNIST.

FEATURE NUMBER	F-TEST		KS-TEST	
	TEST STATISTIC	P-VALUE	SUMMED TEST STATISTIC	SUMMED P- VALUE
1	NAN	NAN	420.83	0.4576
2	NAN	NAN	418.96	0.4576
3	NAN	NAN	417.34	$4.13 \cdot 10^{-4}$
4	NAN	NAN	409.15	$6.39 \cdot 10^{-4}$
⋮	⋮	⋮	⋮	⋮
68	3.17	$7.8 \cdot 10^{-4}$	323.49	0.06
69	2.54	0.0065	323.14	0.99
⋮	⋮	⋮	⋮	⋮
783	0.18	0.996	143.49	8.32
784	0.15	0.998	143.48	9.48
VARIANCE	NaN	NaN	5,038.9	14,094.0

The Ecoli dataset considers measurements of various Ecoli cells relating to different biological aspects [587]. The original dataset contains eight classes, related to the localization site of the Ecoli [587]. This was condensed into four groups (Cytoplasm, Inner Membrane, Periplasm, and Outer Membranes) due to the presence of very small

minority classes. When the KS-test and F-test DRA methods are applied, again one see the recurring issues with p-value but not with test statistic values, Table C-9.

Table C-9: p-values vs Test Statistic for Ecoli.

FEATURE NUMBER	F-TEST		KS-TEST	
	TEST STATISTIC	P-VALUE	SUMMED TEST STATISTIC	SUMMED P- VALUE
1	52.34	$8.30 \cdot 10^{-50}$	1.65	0.039
2	61.94	$2.65 \cdot 10^{-56}$	1.69	0.11
3	109.46	$6.84 \cdot 10^{-82}$	3.59	$1.00 \cdot 10^{-36}$
4	46.58	$1.32 \cdot 10^{-45}$	3.68	$1.07 \cdot 10^{-36}$
5	28.18	$2.76 \cdot 10^{-30}$	1.79	0.11
6	181.38	$1.03 \cdot 10^{-108}$	1.68	0.43
7	93.65	$2.36 \cdot 10^{-74}$	1.78	0.41
VARIANCE	2,700	$1.09 \cdot 10^{-60}$	0.88	0.03

Of particular interest was the generalizability of the benefits of test-statistic feature relevance ranking over p-value for feature relevance ranking. This was demonstrated in all cases except MNIST. This was again due to the representative academic dataset having a machine precision issue when using p-values for feature relevance ranking, but not when using test statistics. Some statistical software truncates p-values at a certain point, e.g. JMP truncates p-values and list them as “<0.0001” [593], to avoid computing infinitesimally small values. While such an approach would avoid presenting and using values beyond machine precision, such approaches are logically also insufficient for feature relevance ranking. No such issues existed with the test statistic values, and only in the Wine dataset were two identical test statistical values computed

for two features using the KS-test; however, this was the only occurrence of this type of problem seen across all of this datasets and does not negate the various obvious issues seen in the p -value rankings.

Throughout all of these academic datasets and the ZigBee RF-DNA dataset, no such issues existed for the test statistic relevance ranking. This both illustrates the generalizability of the results in Section 4.2.1.3 to a wide range of problems and dataset sizes and empirically verifies the recommendation of [365] regarding p -values and feature relevance ranking.

As seen in the MNIST data, KS-test has the benefit that variables consisting of all 0s or identical values can still be examined, while the F-test does not. However, such situations indicate that variables with such conditions will make the data singular or nearly singular, which will preclude further analysis in MDA or other linear classifiers. Nonlinear and ANN based classifiers may still be able to operate on such data, however variables that are identically one value would be necessarily redundant.

APPENDIX D: DRA Method Fusion Classification and Verification Performance Assessments

The chess board is the world, the pieces the phenomena of the universe, the rules of the game are what we call the laws of nature. The player on the other side is hidden from us. All we know is that his play is always fair, just and patient. But, also, that he never overlooks a mistake or makes the smallest allowance for ignorance.

–THOMAS HENRY HUXLEY, 1825 – 1895

By considering the DRA fusion methods in Section 4.2.4 one can determine if fusion of DRA methods offers any performance benefit. MDA/ML models were constructed using the DRA fusion methods and then classification and verification accuracy of each model are presented, respectively, in Table D-1 and Table D-2. Table D-1 shows that DRA fusion methods achieve consistently worse performance than the best result seen in the DRA methods by themselves (presented in the last column of Table D-1). However, while score and rank fusion offer consistently poor performance, concatenation DRA fusion offers performance similar performance to the original DRA methods. Thus concatenation DRA fusion might be viable since it balances the contributions and weaknesses of various methods.

Table D-1: Relative DRA “Gain” (dB) Over Baseline Performance for %C = 90% Classification Accuracy for DRA Fusion Methods. Bold entries denote values within 10% of the Best, and bold entries with light grey shading denote best case performance.

SET		FUSION METHOD			BEST RESULT FROM TABLE IV-6
		Score	Rank	Concatenate	
$N_F=26$	TRAINING	-18.462	–	–13.215	–13.347
	TESTING	-18.393	–	–13.852	–13.817
$N_F=50$	TRAINING	-8.712	–16.972	–9.324	–7.697
	TESTING	-8.513	–17.343	–9.482	–7.967
$N_F=100$	TRAINING	-4.732	–12.532	–4.105	–3.387
	TESTING	-4.643	–12.563	–4.002	–3.407
$N_F=157$	TRAINING	-2.792	–10.822	–2.475	–2.207
	TESTING	-2.683	–10.773	–2.272	–2.357
$N_F=191$	TRAINING	-2.362	–10.152	–2.095	–1.767
	TESTING	-2.303	–10.223	–1.972	–1.917

Table D-2: Device ID Verification Performance for $\%C = 90\%$ at $SNR = 10$ dB: True Verification Rate (TVR) for $N_{Auth} = 4$ Authorized Devices and Rogue Rejection Rate (RRR) For $N_{Auth} \times N_{Rog} = 36$ rogue scenarios. Bold entries denote values within 10% of the Best, and bold entries with light grey shading denote best case performance and.

SET		FUSION METHOD			BEST RESULT FROM TABLE IV-8
		SCORE	RANK	CONCATENATE	
$N_F=10$	AUTHORIZED	0	0	25	50
	ROGUE	19.44	0	38.89	52.78
$N_F=26$	AUTHORIZED	50	0	25	50
	ROGUE	66.67	0	75	80.56
$N_F=50$	AUTHORIZED	50	0	50	75
	ROGUE	88.89	0	86.11	91.67
$N_F=100$	AUTHORIZED	75	0	75	100
	ROGUE	97.22	11.11	94.44	94.44
$N_F=157$	AUTHORIZED	100	25	75	100
	ROGUE	97.22	41.67	94.44	94.44
$N_F=191$	AUTHORIZED	100	50	100	100
	ROGUE	97.22	55.56	97.22	94.44

The verification results from DRA fusion, Table D-2, show a similar deficiency in DRA fusion methods as seen in Table D-1. Again, DRA fusion methods consistently underperform individual DRA methods for verification, particularly at low N_{DRA} . At higher N_{DRA} , e.g. $N_{DRA} = [100, 157, 191]$, DRA fusion methods are seen to achieve comparable or better performance to the individual DRA methods. However, this it should be taken in consideration that the performance differences seen are very slight. Thus DRA fusion methods have limited applicability to RF-DNA classifier model development when compared to using the original DRA methods.

APPENDIX E: Gradient Descent and Derivatives in GLVQ Family Algorithms

...artificial networks need not imitate biology.

–TEUVO KOHONEN, 1934 –

In GLVQ the cost function is no longer the distance measure itself and is now expressed as a function of both a sigmoid, (3.33), and a relative distance measure, (3.34), which is itself a function of both the nearest in-class and out-of-class distances. Overall, these changes complicate the derivation process and the process must be examined closely.

The cost function itself is first examined. Correctly, to compute the first derivative, one must consider that the derivative is with respect to the appropriate PV, w^J or w^L . However, since the in/out-of-class aspect of the PV is not functionally relevant this can be generalized as $df(\mu(x^m))/dw$. First, considering $\partial f(\mu(x^m))/\partial \mu(x^m)$, one must realize that $\mu(x^m)$ is a function within $f(\mu(x^m))$, therefore this can be solved via the chain rule as described in (3.22). With this approach, the gradient of the cost function can be computed as

$$\frac{df(\mu(x^m))}{dw} = \frac{\partial f(\mu(x^m))}{\partial \mu(x^m)} \frac{\partial \mu(x^m)}{\partial w}. \quad (\text{E.1})$$

with

$$\frac{\partial f(\mu(x^m))}{\partial \mu(x^m)} = f'(\mu(x^m))\mu'(x^m) \quad (\text{E.2})$$

where $f'(\mu(x^m)) = -(1 + e^{-\mu(x^m)})^{-2}$ due to the formulation in (3.32)–(3.34) thus yielding the following

$$\frac{\partial f(\mu(x^m))}{\partial \mu(x^m)} = -\left(\frac{1}{1 + e^{-\mu(x^m)}}\right)^2 \frac{\partial}{\partial \mu(x^m)}(1 + e^{-\mu(x^m)}) \quad (\text{E.3})$$

which, because of the expression in (D.2), reduces to

$$\frac{\partial f(\mu(x^m))}{\partial \mu(x^m)} = \left(\frac{1}{1 + e^{-\mu(x^m)}}\right) \left(\frac{e^{-\mu(x^m)}}{1 + e^{-\mu(x^m)}}\right) \quad (\text{E.4})$$

or

$$\frac{\partial f(\mu(x^m))}{\partial \mu(x^m)} = f(\mu(x^m)) (1 - f(\mu(x^m))). \quad (\text{E.5})$$

With a solution to $\partial f(\mu(x^m))/\partial \mu(x^m)$, one must now solve for $\partial \mu(x^m)/\partial w$. Since $\mu(x^m)$ is expressed in the form seen in (3.34), $\partial \mu(x^m)/\partial w$ can be solved via a quotient rule,

$$\partial \left(\frac{u}{v}\right) = \frac{v\partial u - u\partial v}{v^2}, \quad (\text{E.6})$$

where the derivative of both the numerator and denominator must be computed [276]. Per (46), $v = (d^J + d^L)$, $u = (d^J - d^L)$, and $v^2 = (d^J + d^L)^2$, leaving dv and du to be computed. One must realize that dv and du are both a function of the *in*-class or out-of-class, w^J and w^L respectively, PV gradient descents, therefore computing dv and du involves solving four derivatives to yield two equations for the in-class and out-of-class

gradient descents, $\frac{\partial u}{\partial w^J}$ and $\frac{\partial v}{\partial w^J}$ for w^J and $\frac{\partial u}{\partial w^L}$ and $\frac{\partial v}{\partial w^L}$ for w^L respectively. All four derivatives can be generally expressed as

$$\frac{\partial u}{\partial w^{J,L}} = \frac{\partial(d^J \pm d^L)}{\partial w^{J,L}} \quad (\text{E.7})$$

and the derivatives computed via the sum of derivatives rule,

$$\partial(u + v) = \partial u + \partial v. \quad (\text{E.8})$$

For derivatives associated with u , (E.7) can be expressed as

$$\frac{\partial(d^J - d^L)}{\partial w^{J,L}} = \frac{\partial d^J}{\partial w^J} - \frac{\partial d^L}{\partial w^L} \quad (\text{E.9})$$

and similarly for v as

$$\frac{\partial v}{\partial w^{J,L}} = \frac{\partial(d^J + d^L)}{\partial w^{J,L}} = \frac{\partial d^J}{\partial w^J} + \frac{\partial d^L}{\partial w^L}. \quad (\text{E.10})$$

Obviously, depending on whether these derivatives are computed for dw^J or dw^L , one of these components will equal zero and the other will be computed via the derivative of the distance metric. Therefore, the GLVQ gradient derivative formulation can be simplified to the following two general equations, $\partial \left(\frac{u}{v}\right)^J$ and $\partial \left(\frac{u}{v}\right)^L$ which is simplified since $\partial v^J = \partial u^J$ and $\partial v^L = -\partial u^L$,

$$\partial \left(\frac{u}{v}\right)^J = \frac{u\partial u^J - v\partial v^J}{v^2}, \quad (\text{E.11})$$

and

$$\partial \left(\frac{u}{v}\right)^L = \frac{u\partial u^L - v\partial v^L}{v^2}, \quad (\text{E.12})$$

this can further be simplified to:

$$\partial \left(\frac{u}{v} \right)^J = \frac{\partial u^J (u - v)}{v^2} \quad (\text{E.13})$$

and

$$\partial \left(\frac{u}{v} \right)^L = \frac{\partial u^L (u + v)}{v^2}. \quad (\text{E.14})$$

Inserting our expressions for u and v into (E.13) and (E.14) yields,

$$\partial \left(\frac{u}{v} \right)^J = \frac{\partial u^J ((d^J - d^L) - (d^J + d^L))}{(d^J + d^L)^2} = \frac{\partial u^J (-2d^L)}{(d^J + d^L)^2} \quad (\text{E.15})$$

and

$$\partial \left(\frac{u}{v} \right)^L = \frac{\partial u^L ((d^J - d^L) + (d^J + d^L))}{(d^J + d^L)^2} = \frac{\partial u^L (2d^J)}{(d^J + d^L)^2}. \quad (\text{E.16})$$

which provides the negation to make the in-class PV operation move closer and the out-of-class PV move further away. From this formulation, and assuming one doesn't change the cost function itself, to change distance metrics one must merely compute the first derivate of the respective distance metric with respect to both the in-class and out-of-class PV and insert it appropriately. If one has examined changing distance metrics in the LVQ process first, then one only needs to consider the computed first derivative and appropriately add superscripts to designate in-class and out-of-class distance.

For the nominal squared Euclidean distance metric, this is solved via the chain rule and hence all derivatives are multiplied by -1 due to the negative w term. One can then solve (E.15) for ∂u^J

$$\partial u^J = \frac{\partial u}{\partial w^J} = \frac{\partial d^J}{\partial w^J} = 2(x^m - w^J) \cdot -1 = -2(x^m - w^J) \quad (\text{E.17})$$

and for ∂v^J

$$\partial v^J = \frac{\partial v}{\partial w^J} = \frac{\partial d^J}{\partial w^J} = 2(x^m - w^J) \cdot -1 = -2(x^m - w^J). \quad (\text{E.18})$$

Then (E.16) can be solve for ∂u^J

$$\partial u^L = \frac{\partial u}{\partial w^L} = -\frac{\partial d^L}{\partial w^L} = -2(x^m - w^L) \cdot -1 = 2(x^m - w^L) \quad (\text{E.19})$$

and ∂v^J

$$\partial v^L = \frac{\partial v}{\partial w^L} = \frac{\partial d^L}{\partial w^L} = 2(x^m - w^L) \cdot -1 = -2(x^m - w^L). \quad (\text{E.20})$$

To compute the equation for the gradient descent updates, one must place the appropriate components into (E.6) for *in-class* or *out-of-class* gradient descents, w^J and w^L respectively denoted as d^J and d^L , yields

$$\partial \left(\frac{u}{v} \right)^J = \frac{(d^J + d^L)(-2(x^m - w^J)) - (d^J - d^L)(-2(x^m - w^J))}{(d^J + d^L)^2}, \quad (\text{E.21})$$

and

$$\partial \left(\frac{u}{v} \right)^L = \frac{(d^J + d^L)(2(x^m - w^J)) - (d^J - d^L)(-1 * 2(x^m - w^J))}{(d^J + d^L)^2}, \quad (\text{E.22})$$

which can be expressed as

$$\partial \left(\frac{u}{v} \right)^J = \frac{-2(x^m - w^J)((d^J + d^L) - (d^J - d^L))}{(d^J + d^L)^2}, \quad (\text{E.23})$$

and

$$\partial \left(\frac{u}{v} \right)^L = \frac{2(x^m - w^L)((d^J + d^L) + (d^J - d^L))}{(d^J + d^L)^2} \quad (\text{E.24})$$

which further reduces to

$$\partial \left(\frac{u}{v} \right)^J = \frac{-2(x^m - w^J)(2d^L)}{(d^J + d^L)^2} \quad (\text{E.25})$$

and

$$\partial \left(\frac{u}{v} \right)^L = \frac{2(x^m - w^L)(2d^J)}{(d^J + d^L)^2} \quad (\text{E.26})$$

which yields,

$$\partial \left(\frac{u}{v} \right)^J = -\frac{4(x^m - w^J)d^L}{(d^J + d^L)^2} \quad (\text{E.27})$$

and

$$\partial \left(\frac{u}{v} \right)^L = \frac{4(x^m - w^L)d^J}{(d^J + d^L)^2}, \quad (\text{E.28})$$

which is the derivative of the distance used in the quotient rule, within the chain rule.

The gradient descent for GRLVQ type algorithms is then the gradient by chain rule

$$\frac{\partial f(\mu(x^m))}{\partial d^J} \frac{4d^{J,L}}{(d^J + d^L)^2} \quad (\text{E.29})$$

multiplied by the learning rate, $\epsilon(t)$, and a differential shifting,

$$(x^m - w^{J,L}), \quad (\text{E.30})$$

which yields the gradient descent equations in (3.38).

APPENDIX F: Gradient Descent in GRLVQ and GRLVQI Relevance Computation

Those who are good at archery learnt from the bow and not from Yi the Archer. Those who know how to manage boats learnt from the boats and not from Wo. Those who can think learnt from themselves, and not from the Sages.

—ANONYMOUS (T'ANG DYNASTY)¹

For GRLVQ and GRLVQI, the relevance computations and relevance gradient descent must be considered. GRLVQ and GRLVQI extend GLVQ in a similar manner as RLVQ extends LVQ. Thus the PV update in GRLVQ and GRLVQI are consistent with the gradient update in Section 5.2.4, and the relevance computation in GRLVQ and GRLVQI is associated with a gradient descent. As in Section 5.2.2.2(a), this is a function of ψ_q and it would be computed as $\partial f(\mu(x^m))/\partial\psi$, or

$$\frac{\partial f(\mu(x^m))}{\partial\psi} = \frac{\partial f(\mu(x^m))}{\partial\mu(x^m)} \frac{\partial\mu(x^m)}{\partial\psi}, \quad (\text{F.1})$$

with $\frac{\partial f(\mu(x^m))}{\partial\mu(x^m)}$ already solved for the PV update, in (D.2)–(D.5). Therefore, solving (F.1) involves solving $\frac{\partial\mu(x^m)}{\partial\psi}$, which involves a logically similar approach to solving for $\partial\mu(x^m)/\partial w$, via the quotient rule in (E.6), only with $v = (d^J + d^L)$, $u = (d^J - d^L)$, and $v^2 = (d^J + d^L)^2$, for

$$\frac{\partial u}{\partial\psi} = \frac{\partial(d^J - d^L)}{\partial\psi} = \frac{\partial d^J}{\partial\psi} - \frac{\partial d^L}{\partial\psi} \quad (\text{F.2})$$

and for v

¹From the 8th Century Taoist book *Kuan Yin Tzu*

$$\frac{\partial v}{\partial \psi} = \frac{\partial(d^J + d^L)}{\partial \psi} = \frac{\partial d^J}{\partial \psi} + \frac{\partial d^L}{\partial \psi}. \quad (\text{F.3})$$

For the nominal squared Euclidean distance equation, components of (F.2) and (F.3) can be solved as

$$\frac{\partial d^J}{\partial \psi} = \psi \cdot 0 + 1 \cdot (x_{iq}(t) - w_{nq}(t))^2 = (x_{iq}(t) - w_{nq}(t))^2 \quad (\text{F.4})$$

and

$$\frac{\partial d^L}{\partial \psi} = \psi \cdot 0 + 1 \cdot (x_{iq}(t) - w_{nq}(t))^2 = (x_{iq}(t) - w_{nq}(t))^2. \quad (\text{F.5})$$

Since

$$\frac{\partial d^J}{\partial \psi} = \frac{\partial d^L}{\partial \psi}, \quad (\text{F.6})$$

and

$$\frac{\partial u}{\partial \psi} = 0, \quad (\text{F.7})$$

then, for dv , we can arrive at the solution:

$$\frac{\partial v}{\partial \psi} = 2(x_{iq}(t) - w_{nq}(t))^2. \quad (\text{F.8})$$

Putting this together and solving for $\partial \left(\frac{u}{v} \right)$ via the quotient rule yields the following,

$$\partial \left(\frac{u}{v} \right) = - \frac{2(d^J - d^L)(x_{iq}(t) - w_{nq}(t))^2}{(d^J + d^L)^2} \quad (\text{F.9})$$

which, yields a PV update,

$$\begin{aligned} \psi_q(t+1) = \psi_q(t) \\ - \epsilon(t) f' |_{\mu(x^m)} \left(- \frac{2(d^J - d^L)(x_{iq}(t) - w_{nq}(t))^2}{(d^J + d^L)^2} \right) \end{aligned} \quad (\text{F.10})$$

which is equivalent to the GRLVQ update in (3.37) prior to being multiplied and written out.

Because the improvements in GRLVQI consist of scalar learning rates and criteria outside the distance metric and cost function, the PV update process is not different from that of GRLVQ. Therefore the PV update process presented for GRLVQ and GLVQ can be directly applied.

APPENDIX G: Cost Function Extensions for the GLVQ Family of Algorithms

A 'simple analysis' can be harder than it looks...

–CHRISTOPHER CHATFIELD

From Sections 5.2.2.2(b) and 5.2.2.4 it is known that not all derivatives need to be recomputed. Since changing $\mu(x^m)$ does not change the cost function expression in (3.34), then only the derivative for the second part of (E.1), $\partial\mu(x^m)/\partial w$, must be recomputed. Again, following the quotient rule in (E.6), we determine the respective quantities for (5.9) as $v = (d^J)^2 + (d^L)^2$, $u = (d^J)^2 - (d^L)^2$, and $v^2 = ((d^J)^2 + (d^L)^2)^2$, with again dv and du to be computed for the respective in/out of class PVs. Then the process in Section 5.2.2.2(b) is repeated to arrive at new PV update rules. Again, four derivatives to yield two equations for the in-class and out-of-class gradient descents, $\frac{\partial u}{\partial w^J}$ and $\frac{\partial v}{\partial w^J}$ for w^J and $\frac{\partial u}{\partial w^L}$ and $\frac{\partial v}{\partial w^L}$ for w^L respectively. Similar to the general derivative in (E.7), all four derivatives can be generally expressed as

$$\frac{\partial u}{\partial w^{J,L}} = \frac{\partial((d^J)^2 - (d^L)^2)}{\partial w^{J,L}}, \quad (\text{G.1})$$

with the derivative for u expressed as

$$\frac{\partial((d^J)^2 - (d^L)^2)}{\partial w^{J,L}} = \frac{\partial(d^J)^2}{\partial w^J} - \frac{\partial(d^L)^2}{\partial w^L} \quad (\text{G.2})$$

and the derivative for v expressed as

$$\frac{\partial v}{\partial w^{J,L}} = \frac{\partial((d^J)^2 + (d^L)^2)}{\partial w^{J,L}} = \frac{\partial(d^J)^2}{\partial w^J} + \frac{\partial(d^L)^2}{\partial w^L}. \quad (\text{G.3})$$

Consistent with 5.2.4, if dw^J or dw^L is of interest one of these components will equal zero and the other will be computed via the derivative of the distance metric. Since the GLVQ gradient descent formulation has not been altered, we can use the quotient rule derivatives in (E.13) and (E.14) to insert our expressions for u and v into (E.13) and (D.14) yields,

$$\partial \left(\frac{u}{v} \right)^J = \frac{\partial u^J((d^J)^2 - (d^L)^2) - ((d^J)^2 + (d^L)^2)}{((d^J)^2 + (d^L)^2)^2} = \frac{\partial u^J(-2(d^L)^2)}{((d^J)^2 + (d^L)^2)^2} \quad (\text{G.4})$$

and

$$\partial \left(\frac{u}{v} \right)^L = \frac{\partial u^L((d^J)^2 - (d^L)^2) + ((d^J)^2 + (d^L)^2)}{((d^J)^2 + (d^L)^2)^2} = \frac{\partial u^L(2(d^J)^2)}{((d^J)^2 + (d^L)^2)^2}. \quad (\text{G.5})$$

Next, one can then solve (E.15) for ∂u^J where the differential shifting for the ∂u^L and ∂u^J ; firstly, we compute

$$\begin{aligned} \partial u^J &= \frac{\partial u}{\partial w^J} = \frac{\partial((d^J)^2)}{\partial w^J} = \frac{\partial(x^m - w^J)^4}{\partial w^J} \\ &= 4(x^m - w^J)^3 \cdot -1 = -4(x^m - w^J)^3 \end{aligned} \quad (\text{G.6})$$

and for ∂v^J

$$\begin{aligned} \partial v^J &= \frac{\partial v}{\partial w^J} = \frac{\partial(d^L)^2}{\partial w^J} = 4(x^m - w^J)^3 \cdot -1 \\ &= -4(x^m - w^J)^3. \end{aligned} \quad (\text{G.7})$$

Then (E.16) can be solved for ∂u^J

$$\begin{aligned} \partial u^L &= \frac{\partial u}{\partial w^L} = -\frac{\partial(d^J)^2}{\partial w^L} = -4(x^m - w^L)^3 \cdot -1 \\ &= 4(x^m - w^L)^3 \end{aligned} \quad (\text{G.8})$$

and dv^J

$$\begin{aligned}\partial v^L &= \frac{\partial v}{\partial w^L} = \frac{\partial (d^L)^2}{\partial w^L} = 4(x^m - w^L)^3 \cdot -1 \\ &= -4(x^m - w^L)^3.\end{aligned}\tag{G.9}$$

Assembling all of these components, one can fully extend to a PV update equation

$$\begin{aligned}w^J(t+1) &= w^J(t) + \frac{8\epsilon(t)(\partial f / \partial \mu(x^m))d^L}{(d^J + d^L)^2}(x^m - w^J)^3 \\ w^K(t+1) &= w^K(t) - \frac{8\epsilon(t)(\partial f / \partial \mu(x^m))d^J}{(d^J + d^L)^2}(x^m - w^K)^3.\end{aligned}\tag{G.10}$$

which differs from the PV updates in (3.35) only by the scalar multiplier and the squared terms in the relative distance difference equations.

APPENDIX H: Relevance Derivatives for GRLVQI

Remember it takes time, patience, critical practice and knowledge to learn any art or craft.

–LLOYD REYNOLDS, 1902-1978

As previously noted in Sections 3.3.1.4, 3.3.1.6, 5.2.2.2(a), and 5.2.2.2(b), relevance learning in LVQ algorithms involves a further gradient descent operation. Therefore, when considering alternative distance measures for GRLVQ and GRLVQI, the relevance computations and relevance gradient descent must be considered. As in RLVQ, the relevance computation in GRLVQ and GRLVQI is associated with a gradient descent; therefore to compute the GRLVQ and GRLVQI update equations, we must revisit the gradient descent computations in Section 5.2.2.2(b) using the gradient update in (G.10) and relative distance difference (5.9). Again, as in Section 5.2.2.2(a), if this is a function of the ψ_q , then it would be computed as $\partial f(\mu(x^m))/\partial\psi$, or

$$\frac{\partial f(\mu(x^m))}{\partial w} = \frac{\partial f(\mu(x^m))}{\partial \mu(x^m)} \frac{\partial \mu(x^m)}{\partial \psi}. \quad (\text{H.1})$$

with $\frac{\partial f(\mu(x^m))}{\partial \mu(x^m)}$ already solved for the PV update, in (E.2) to (E.5). Therefore, solving (F.1) involves solving $\frac{\partial \mu(x^m)}{\partial \psi}$, which involves a logically similar approach to solving for $\partial \mu(x^m)/\partial w$, via the quotient rule in (E.6), only with $v = (d^J + d^L)$, $u = (d^J - d^L)$, and $v^2 = (d^J + d^L)^2$, for

$$\frac{\partial u}{\partial \psi} = \frac{\partial(d^J - d^L)}{\partial \psi} = \frac{\partial d^J}{\partial \psi} - \frac{\partial d^L}{\partial \psi} \quad (\text{H.2})$$

and for v

$$\frac{\partial v}{\partial \psi} = \frac{\partial(d^J + d^L)}{\partial \psi} = \frac{\partial d^J}{\partial \psi} + \frac{\partial d^L}{\partial \psi}. \quad (\text{H.3})$$

For the nominal squared Euclidean distance equation, components of (F.2) and (F.3) can be solved as

$$\frac{\partial d^J}{\partial \psi} = \psi \cdot 0 + 1 \cdot (x_{iq}(t) - w_{nq}(t))^2 = (x_{iq}(t) - w_{nq}(t))^2 \quad (\text{H.4})$$

and

$$\frac{\partial d^L}{\partial \psi} = \psi \cdot 0 + 1 \cdot (x_{iq}(t) - w_{nq}(t))^2 = (x_{iq}(t) - w_{nq}(t))^2. \quad (\text{H.5})$$

Since,

$$\frac{\partial d^J}{\partial \psi} = \frac{\partial d^L}{\partial \psi} \quad (\text{H.6})$$

and

$$\frac{\partial u}{\partial \psi} = 0, \quad (\text{H.7})$$

then, for dv , we can arrive at the solution:

$$\frac{\partial v}{\partial \psi} = 2(x_{iq}(t) - w_{nq}(t))^2. \quad (\text{H.8})$$

Putting this together and solving for $\partial \left(\frac{u}{v} \right)$ via the quotient rule yields the following,

$$\partial \left(\frac{u}{v} \right) = - \frac{2(d^J - d^L)(x_{iq}(t) - w_{nq}(t))^2}{(d^J + d^L)^2} \quad (\text{H.9})$$

which, yields a relevance update,

$$\begin{aligned} \psi_q(t+1) = & \psi_q(t) \\ & - \epsilon(t) f'|_{\mu(x^m)} \left(- \frac{2(d^J - d^L)(x_{iq}(t) - w_{nq}(t))^2}{(d^J + d^L)^2} \right) \end{aligned} \quad (\text{H.10})$$

which is equivalent to the GRLVQ update in (3.38) prior to being multiplied and written out.

APPENDIX I: Review of Distance Measures

One accurate measurement is worth a thousand expert opinions.

–ADMIRAL GRACE HOPPER, 1906 – 1992

Various distance metrics exist, with some literature offering comparisons. Jones and Furnas [594] compared the inner product, cosine measure, pseudo-cosine measure, dice measure, produce-moment correlation and covariance, and overlap measure. Zhang and Korfhage [595] offered further analysis of the cosine measure. Both Cha [283] and McGill et al. [596] produced a review of distance measures, in general these reviews overlapped each other except McGill included binary distance metrics. From these sources, the following review of distance metrics was produced; below, \mathbf{P} and \mathbf{Q} are considered to be two different vectors of equal length, n .

Cha [283] considers the Minkowski family to have four measures, all of which are special cases of the general Minkowski distance,

$$d_{Mk} = \sqrt[p]{\sum_{i=1}^n |P_i - Q_i|^p}, \quad (\text{I.1})$$

which, for $p = 2$, is the Euclidean L_2 distance

$$d_{Eucl} = \sqrt{\sum_{i=1}^n (P_i - Q_i)^2}, \quad (\text{I.2})$$

City Block, for $p = 1$,

$$d_{city} = \sum_{i=1}^n |P_i - Q_i|, \quad (I.3)$$

and Chebyshev, for $p = \infty$,

$$d_{cheb} = \max_i |P_i - Q_i|. \quad (I.4)$$

The L_I family of measures includes many measures, which are variations on the City Block, L_I , measure through division or scaling. Due to the various methods involves, the L_I family deserves some consideration. The Sorensen measure [284],

$$d_{sor} = \frac{\sum_{i=1}^n |P_i - Q_i|}{\sum_{i=1}^n (P_i + Q_i)} \quad (I.5)$$

is typical of the L_I [283]. The *Gower* distance metric is merely a scaling of d_{city} by a scalar and is hence differs from d_{city} by only a magnitude [283], for this reason it is not examined herein. The Soergel, d_{sg} , and Kulczynski, d_{kd} , measures are similar approaches are variants of Sorensen with the maximum, $\sum_{i=1}^n \max(P_i, Q_i)$, or minimum, $\sum_{i=1}^n \min(P_i, Q_i)$, in the denominator, respectively [283]. As noted by Cha [283], the Canberra measure differs from Sorensen through normalizing the absolute difference of the individual level,

$$d_{can} = \sum_{i=1}^n \frac{|P_i - Q_i|}{P_i + Q_i}. \quad (I.6)$$

The Lorentzian measure,

$$d_{Lor} = \sum_{i=1}^n \ln(1 + |P_i - Q_i|) \quad (I.7)$$

applies the natural logarithm to the City Block measure, with the addition of 1 is used to avoid computing the logarithm of zero [283].

Many of the intersection family of distance measures are L_1 based and identical to an L_1 distance measure through a division or subtraction. Examples include the Ruzicka measure,

$$d_{Ruz} = \frac{\sum_{i=1}^n \min(P_i, Q_i)}{\sum_{i=1}^n \max(P_i, Q_i)}, \quad (I.8)$$

which appears different, but is essentially d_{sg}/d_{kd} . This is similar for the Kulczynski measure, which is $1/d_{kd}$, the Intersection measure, which is $\frac{1}{2}d_{City}$, and the Czenkanowski measure, which is identical to Sorensen, and Motyka, which is $\frac{1}{2}d_{Sor}$ [283]. However, some other Intersection family measures are different enough to warrant evaluation, including Wave Hedges,

$$d_{Wave} = \sum_{i=1}^n \frac{|P_i - Q_i|}{\max(P_i, Q_i)}. \quad (I.9)$$

and Tanimoto,

$$d_{Tani} = \frac{\sum_{i=1}^n (\max(P_i, Q_i) - \min(P_i, Q_i))}{\sum_{i=1}^n \max(P_i, Q_i)}. \quad (I.10)$$

The Inner Product family are a group of measures that involve computing the inner product, $P \cdot Q$, of vectors in question [283]. The inner product measure,

$$d_{IP} = P \cdot Q = \sum_{i=1}^n P_i Q_i. \quad (I.11)$$

reflects this. Many of the measures in this family include the inner product computation along with other components. The Harmonic mean scales d_{IP} ,

$$d_{HM} = \sum_{i=1}^n \frac{P_i Q_i}{P_i + Q_i} \quad (I.12)$$

Cha [283] presents the cosine measure as the inner product metric with a further scaling in the denominator,

$$d_{COS} = \frac{\sum_{i=1}^n P_i Q_i}{\sqrt{\sum_{i=1}^n P_i^2} \sqrt{\sum_{i=1}^n Q_i^2}}. \quad (I.13)$$

A variant on the cosine measure is the pseudo-cosine measure

$$d_{PCOS} = \frac{\sum_{i=1}^n P_i Q_i}{\sum_{i=1}^n P_i \sum_{i=1}^n Q_i}. \quad (I.14)$$

which differs from the cosine measure in how it measures vector length [594]. Cha [283]

also presents the Kumar-Hassebrook metric, another extension of the cosine measure,

$$d_{KumarH} = \frac{\sum_{i=1}^n P_i Q_i}{\sum_{i=1}^n P_i^2 + \sum_{i=1}^n Q_i^2 - \sum_{i=1}^n P_i Q_i}. \quad (I.15)$$

Jaccard,

$$d_{Jac} = \frac{\sum_{i=1}^n (P_i - Q_i)^2}{\sum_{i=1}^n P_i^2 + \sum_{i=1}^n Q_i^2 - \sum_{i=1}^n P_i Q_i} \quad (I.16)$$

and Dice [597],

$$d_{Dice} = \frac{\sum_{i=1}^n (P_i - Q_i)^2}{\sum_{i=1}^n P_i^2 + \sum_{i=1}^n Q_i^2} \quad (I.17)$$

measures are also related to the inner product family [283].

The Fidelity family appears similar to the Inner Product family; however, these include natural logarithms and square roots in the distance computations. While these could sufficiently alter the distance metrics, these could also present problems when negative values are introduced and thus cause imaginary numbers to be computed. Therefore these will not be considered, but are presented for completeness. The basic measure in this family, Fidelity is the Inner Product distance with a square-root,

$$d_{Fid} = \sum_{i=1}^n \sqrt{P_i Q_i}. \quad (I.18)$$

Bhattacharyya is an Fidelity family type of measure and is the natural log of d_{Fid} ,

$$d_{Bhat} = -\ln \sum_{i=1}^n \sqrt{P_i Q_i}. \quad (I.19)$$

Hellinger involves a scaling of inner product,

$$d_{Hell} = 2 \sqrt{1 - \sum_{i=1}^n \sqrt{P_i Q_i}}. \quad (I.20)$$

Matusita involves a further scaling,

$$d_{Mat} = \sqrt{2 - 2 \sum_{i=1}^n \sqrt{P_i Q_i}}. \quad (I.21)$$

However, Squared-Chord,

$$d_{SC} = \sum_{i=1}^n (\sqrt{P_i} - \sqrt{Q_i})^2. \quad (I.22)$$

offers a variation on the fidelity measure and appears identical to d_{Fid} by an offset,

$$1 - d_{SC} = 2 \sum_{i=1}^n \sqrt{P_i Q_i} - 1 = 2d_{Fid} - 1 \text{ [283].}$$

The Squared L_2 family offers squared variations on Euclidean distance, including the squared Euclidean distance of (1), in addition to other variations. These variations could cause metrics to produce different results, hence some should be investigated. The Pearson χ^2 and Neyman χ^2 metrics are similar and differ in the denominator,

$$d_{P\chi^2} = \sum_{i=1}^n \frac{(P_i - Q_i)^2}{Q_i} \quad (I.23)$$

and

$$d_{N\chi^2} = \sum_{i=1}^n \frac{(P_i - Q_i)^2}{P_i} \quad (I.24)$$

respectively [283]. The Squared χ^2 further extends these,

$$d_{S\chi^2} = \sum_{i=1}^n \frac{(P_i - Q_i)^2}{P_i + Q_i} \quad (I.25)$$

and the probabilistic symmetric χ^2 measure is $2d_{S\chi^2}$ [283]. The divergence measure,

$$d_{Div} = \sum_{i=1}^n \frac{(P_i - Q_i)^2}{(P_i + Q_i)^2} \quad (I.26)$$

further extends $d_{S\chi^2}$ [283]. Clark,

$$d_{Clark} = \sqrt{\sum_{i=1}^n \left(\frac{|P_i - Q_i|}{P_i + Q_i} \right)^2} \quad (I.27)$$

and additive symmetric χ^2

$$d_{AS\chi^2} = \sum_{i=1}^n \frac{(P_i - Q_i)^2 (P_i + Q_i)}{P_i Q_i} \quad (I.28)$$

further complete the squared L_2 family [283].

Shannon's entropy family includes additional metrics not encompassed in the other families, including Kullback-Leibler,

$$d_{KL} = \sum_{i=1}^n P_i \ln \frac{P_i}{Q_i} \quad (I.29)$$

Jeffreys,

$$d_{Jeff} = \sum_{i=1}^n (P_i - Q_i) \ln \frac{P_i}{Q_i} \quad (I.30)$$

K divergence,

$$d_{Kd} = \sum_{i=1}^n P_i \ln \frac{2P_i}{P_i + Q_i} \quad (I.31)$$

Topsoe

$$d_{top} = \sum_{i=1}^n \left(P_i \ln \left(\frac{2P_i}{P_i + Q_i} \right) + Q_i \ln \left(\frac{2Q_i}{P_i + Q_i} \right) \right) \quad (\text{I.32})$$

Jensen-Shannon,

$$d_{JS} = \frac{1}{2} \left[\sum_{i=1}^n \left(P_i \ln \left(\frac{2P_i}{P_i + Q_i} \right) \right) + \sum_{i=1}^n \left(Q_i \ln \left(\frac{2Q_i}{P_i + Q_i} \right) \right) \right] \quad (\text{I.33})$$

and Jensen difference,

$$d_{Ja} = \sum_{i=1}^n \left[\frac{P_i \ln P_i + Q_i \ln Q_i}{2} - \frac{P_i + Q_i}{2} \ln \frac{P_i + Q_i}{2} \right]. \quad (\text{I.34})$$

Cha [283] also presents a family of combinations, distance measures incorporating concepts and parts of multiple measures. This family includes Taneja,

$$d_{Tan} = \sum_{i=1}^n \left[\frac{P_i + Q_i}{2} \ln \frac{P_i + Q_i}{2\sqrt{P_i Q_i}} \right], \quad (\text{I.35})$$

Kumar-Johnson,

$$d_{KJ} = \sum_{i=1}^n \left(\frac{(P_i^2 - Q_i^2)^2}{2(P_i Q_i)^{3/2}} \right), \quad (\text{I.36})$$

and the average of L_p for $p = [1, \infty]$,

$$d_{AVG} = \frac{1}{2} \sum_{i=1}^n |P_i - Q_i| + \max_i |P_i - Q_i|. \quad (\text{I.37})$$

A further group of distance measures, termed *vicissitude*, includes additional variations of other metrics. This family includes Vicis-Wave Hedges,

$$d_{V-Wave} = \sum_{i=1}^n \frac{|P_i - Q_i|}{\min(P_i, Q_i)}, \quad (I.38)$$

three variations of Symmetric χ^2 which differ from the Squared L_2 family in the denominator with the denominator of d_{Div} replaced with either $\min(P_i, Q_i)$, $\min(P_i, Q_i)^2$, or $\max(P_i, Q_i)$ [283]. The final mentioned vicissitude metrics include max-symmetric χ^2 ,

$$d_{maxSym\chi^2} = \max\left(\sum_{i=1}^n \frac{(P_i - Q_i)^2}{P_i}, \sum_{i=1}^n \frac{(P_i - Q_i)^2}{Q_i}\right) \quad (I.39)$$

and min-symmetric χ^2 ,

$$d_{minSym\chi^2} = \min\left(\sum_{i=1}^n \frac{(P_i - Q_i)^2}{P_i}, \sum_{i=1}^n \frac{(P_i - Q_i)^2}{Q_i}\right). \quad (I.40)$$

Although not listed in Cha's review, Jones and Furnas [594] also present the following equations for covariance metric,

$$d_{Cov} = \sum_{i=1}^n (P_i - \bar{P})(Q_i - \bar{Q}), \quad (I.41)$$

with \bar{P} and \bar{Q} representing the means of P and Q , and the correlation,

$$d_{Corr} = \frac{\sum_{i=1}^n (P_i - \bar{P})(Q_i - \bar{Q})}{\sqrt{\sum_{i=1}^n (P_i - \bar{P})^2} \sqrt{\sum_{i=1}^n (Q_i - \bar{Q})^2}}, \quad (I.42)$$

distance metric [594]. Additionally, the Mahalanobis statistical distance metric was covered in these reviews, but could be useful. The nominal Mahalanobis distance equation is

$$d_{Mahal} = \sqrt{(P_i - \bar{P})' S^{-1} (P_i - \bar{P})}, \quad (\text{I.43})$$

where S is the data covariance matrix [598]. Mahalanobis distance can be extended to a similarity between two vectors through

$$d_{Mahal(x,y)} = \sqrt{(P_i - Q_i)' S^{-1} (P_i - Q_i)}, \quad (\text{I.44})$$

where S is a pooled covariance matrix. For use herein, squaring (I.44) would be more practical to remove the square root for derivation simplicity.

APPENDIX J: Derivatives and Prototype Vectors Updates for Selected Distance Metrics

There is a measure in all things.

–HORACE, 65BC – 8BC

In this appendix, derivatives for the distance measures selected in Section 5.3.1 are formulated. Derivatives for relevance measures discussed in Section 5.3.4 are also considered as needed here. Per the formulation of the cost functions in LVQ algorithms, derivatives of distance measures and metrics are made with respect to the PV, w , or for the relevance vector, ψ , when relevance components of LVQ algorithms are being considered.

7.1 Cosine

If one considers that the denominator of the cosine measure in (I.13) is a scalar, then we can consider the cosine measure as

$$d_{cos} = \sum_{i=1}^{N_F} \frac{x_i w_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n w_i^2}}, \quad (\text{J.1})$$

where the derivative can then be computed via the quotient rule, (E.6), with $u = x_i w_i$,

$v = \sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n w_i^2}$, and then for the derivative with respect to w : $du = x_i$ and

$dv = \frac{\sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{\sum_{i=1}^n w_i^2}} w_i$. Therefore the derivative via the quotient rule is,

$$\frac{\partial d_{cos}}{\partial w} = \sum_{i=1}^{N_F} \frac{x_i \sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n w_i^2} - x_i w_i^2 \sqrt{\sum_{i=1}^n x_i^2} / \sqrt{\sum_{i=1}^n w_i^2}}{\sum_{i=1}^n x_i^2 \sum_{i=1}^n w_i^2}, \quad (J.2)$$

The Cosine distance measure with relevance learning can be formulated a

$$d_{cos,\psi} = \sum_{i=1}^{N_F} \frac{\psi_i x_i w_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n w_i^2}}. \quad (J.3)$$

Per the quotient rule, (E.6), with $u = \psi_i x_i w_i$, $v = \sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n w_i^2}$, and the then for the derivative with respect to ψ : $du = x_i w_i$ and $dv = 0$, then

$$\frac{\partial d_{cos,\psi}}{\partial \psi} = \sum_{i=1}^{N_F} \frac{x_i w_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n w_i^2}}. \quad (J.4)$$

7.2 Sorensen and Canberra

Sorensen and Canberra are similar expressions. Considering the prototype vectors and exemplar data, Sorensen, from (I.5), is defined as

$$d_{Sor} = \frac{\sum_{i=1}^{N_F} x_i - w_i}{\sum_{i=1}^{N_F} x_i + w_i} \quad (J.5)$$

and Canberra, from (I.6), is defined as

$$d_{Can} = \sum_{i=1}^{N_F} \frac{x_i - w_i}{x_i + w_i} \quad (J.6)$$

with the underlying difference being that Sorensen considers a ratio of sums whereas Canberra considers a sum of ratios. However, while the distance measures produce different distances (which were uncorrelated per the discussion in), both have similar derivations with respect to $\partial/\partial w$. For both Sorensen and Canberra $u = x_i - w_i$, $v = x_i +$

w_i , and the then for the derivative with respect to w : $du = -1$ and $dv = 1$. Therefore the derivatives via the quotient rule are

$$\frac{\partial d_{Sor}}{\partial w} = \frac{\sum_{i=1}^{N_F} -2x_i}{\sum_{i=1}^{N_F} (x + w)^2} \quad (J.7)$$

and

$$\frac{\partial d_{Can}}{\partial w} = \sum_{i=1}^{N_F} \frac{-2x_i}{(x + w)^2}. \quad (J.8)$$

Due to both Sorensen offering consistent, albeit slightly less, performance than Canberra in LVQ and the relative difficulty of introducing a relevance term into the Sorensen expression, only Canberra was further considered for RLVQ, GLVQ, GRLVQ, and GRLVQI. To implement relevance learning, the relevance must be added so that it multiplies to each feature

$$d_{Can,\psi} = \sum_{i=1}^{N_F} \psi_i \frac{x_i - w_i}{x_i + w_i} \quad (J.9)$$

which means $u = \psi_i(x_i - w_i)$, $v = x_i + w_i$, and the then for the derivative with respect to ψ : $du = (x_i - w_i)$, and $dv = 0$. The resulting derivative is therefore,

$$\frac{\partial d_{Can,\psi}}{\partial \psi} = \sum_{i=1}^{N_F} \frac{x_i - w_i}{x_i + w_i}. \quad (J.10)$$

7.3 Pseudo-Cosine

Considering the prototype vectors and exemplar data, the Pseudo Cosine measure of (I.14) becomes

$$d_{pcos} = \sum_{i=1}^{N_F} \frac{x_i w_i}{\sum_{i=1}^{N_F} x_i \sum_{i=1}^{N_F} w_i} \quad (\text{J.11})$$

the derivative can then be computed via the quotient rule, (E.6), can be used to compute the derivative, with $u = x_i w_i$, $v = \sum_{i=1}^{N_F} x_i \sum_{i=1}^{N_F} w_i$, and the then for the derivative with respect to w : $du = x_i$ and $dv = \sum_{i=1}^{N_F} x_i$. Therefore the derivative via the quotient rule is,

$$\frac{\partial d_{pcos}}{\partial w} = \sum_{i=1}^{N_F} \frac{x_i \sum_{i=1}^{N_F} x_i \sum_{i=1}^{N_F} w_i - x_i w_i \sum_{i=1}^{N_F} x_i}{\left(\sum_{i=1}^{N_F} x_i \sum_{i=1}^{N_F} w_i\right)^2}. \quad (\text{J.12})$$

7.4 Pearson χ^2

Considering the prototype vectors and exemplar data, the Pearson χ^2 measure of (I.23) becomes

$$d_{p\chi^2} = \sum_{i=1}^{N_F} \frac{(x_i - w_i)^2}{w_i} \quad (\text{J.13})$$

the derivative can then be computed via the quotient rule, (E.6), can be used to compute the derivative, with $u = (x_i - w_i)^2$, $v = w_i$, and the then for the derivative with respect to w : $du = -2(x_i - w_i)$ and $dv = 1$. Therefore the derivative via the quotient rule is,

$$\frac{\partial d_{p\chi^2}}{\partial w} = \sum_{i=1}^{N_F} \frac{-2x_i(x_i - w_i) - (x_i - w_i)^2}{w_i^2} \quad (\text{J.14})$$

7.5 Neyman χ^2

Considering the prototype vectors and exemplar data, the Neyman χ^2 measure of (I.24) becomes

$$d_{N\chi^2} = \sum_{i=1}^{N_F} \frac{(x_i - w_i)^2}{x_i} \quad (\text{J.15})$$

the derivative can then be computed via the quotient rule, (E.6), can be used to compute the derivative, with $u = (x_i - w_i)^2$, $v = x_i$, and the then for the derivative with respect to w : $ddu = -2(x_i - w_i)$ and $dv = 0$. Therefore the derivative via the quotient rule is,

$$\frac{\partial d_{N\chi^2}}{\partial w} = \sum_{i=1}^{N_F} \frac{-2x_i(x_i - w_i)}{x_i^2}. \quad (\text{J.16})$$

7.6 Additive Symmetry

Considering the prototype vectors and exemplar data, the Additive Symmetry χ^2 measure of (I.28) becomes

$$d_{AS\chi^2} = \sum_{i=1}^{N_F} \frac{(x_i - w_i)^2(x_i - w_i)}{x_i w_i} \quad (\text{J.17})$$

the derivative can then be computed via the quotient rule, (E.6), can be used to compute the derivative, with $u = (x_i - w_i)^2(x_i - w_i)$, $v = x_i w_i$, and the then for the derivative with respect to w : $du = -3w_i^2 - 2x_i w_i + x_i^2$ and $dv = x_i$. Therefore the derivative via the quotient rule is,

$$\begin{aligned} & \frac{\partial d_{AS\chi^2}}{\partial \mathbf{w}} \\ &= \sum_{i=1}^{N_F} \frac{[x_i w_i (-3w_i^2 - 2x_i w_i + x_i^2) - x_i (x_i - w_i)^2 (x_i - w_i)]}{(x_i w_i)^2}. \end{aligned} \quad (\text{J.18})$$

7.7 Covariance

The covariance measure, (I.41), involves determining the means of both the PVs and data. In matrix notation one can express (I.41) as

$$\mathbf{d}_{\text{COV}} = \left(\mathbf{x} - \frac{\mathbf{1}\mathbf{1}'\mathbf{x}}{n} \right)' \left(\mathbf{w} - \frac{\mathbf{1}\mathbf{1}'\mathbf{w}}{n} \right) = \left(\mathbf{x}' - \frac{\mathbf{x}'\mathbf{1}\mathbf{1}'}{n} \right) \left(\mathbf{w} - \frac{\mathbf{1}\mathbf{1}'\mathbf{w}}{n} \right) \quad (\text{J.19})$$

multiplying expression yields,

$$\mathbf{d}_{\text{COV}} = \mathbf{x}'\mathbf{w} - \frac{\mathbf{x}'\mathbf{1}\mathbf{1}'\mathbf{w}}{n} - \frac{\mathbf{x}'\mathbf{1}\mathbf{1}'\mathbf{w}}{n} + \frac{\mathbf{x}'\mathbf{1}\mathbf{1}'\mathbf{1}\mathbf{1}'\mathbf{w}}{n^2} \quad (\text{J.20})$$

Taking the derivative of this expression yields,

$$\frac{\partial \mathbf{d}_{\text{COV}}}{\partial \mathbf{w}} = \mathbf{x}' - \frac{\mathbf{x}'\mathbf{1}\mathbf{1}'}{n} - \frac{\mathbf{x}'\mathbf{1}\mathbf{1}'}{n} + \frac{\mathbf{x}'\mathbf{1}\mathbf{1}'\mathbf{1}\mathbf{1}'}{n^2}, \quad (\text{J.21})$$

which can be simplified algebraically to

$$\frac{\partial \mathbf{d}_{\text{COV}}}{\partial \mathbf{w}} = \mathbf{x}' \left(\mathbf{I} - \frac{\mathbf{J}}{n} \right), \quad (\text{J.22})$$

where \mathbf{I} is an identity matrix and \mathbf{J} is a matrix of ones.

7.8 Squared Mahalanobis

As illustrated in Section 5.3.3.1, Mahalanobis distance and squared Mahalanobis distance are perfectly correlated. Therefore, for use herein, squaring (I.44) was assumed to be more practical to remove the square root for simplicity in derivations. The

covariance \mathbf{S}^{-1} is assumed to be the covariance of the data. In matrix notation, the squared form of (I.44) can be expressed as:

$$d_{Mahal(x,y)} = (\mathbf{x} - \mathbf{w})' \mathbf{S}^{-1} (\mathbf{x} - \mathbf{w}), \quad (\text{J.23})$$

which can be expressed as

$$d_{Mahal(x,y)} = (\mathbf{x}' - \mathbf{w}') \mathbf{S}^{-1} (\mathbf{x} - \mathbf{w}). \quad (\text{J.24})$$

One can now appropriately distribute the covariance matrix,

$$d_{Mahal(x,y)} = (\mathbf{x}' \mathbf{S}^{-1} - \mathbf{w}' \mathbf{S}^{-1}) (\mathbf{x} - \mathbf{w}). \quad (\text{J.25})$$

which expands to

$$d_{Mahal(x,y)} = \mathbf{x}' \mathbf{S}^{-1} \mathbf{x} - \mathbf{x}' \mathbf{S}^{-1} \mathbf{w} - \mathbf{w}' \mathbf{S}^{-1} \mathbf{x} + \mathbf{w}' \mathbf{S}^{-1} \mathbf{w}. \quad (\text{J.26})$$

which has the first derivative

$$\frac{\partial d_{Mahal(x,y)}}{\partial \mathbf{w}} = -2 \mathbf{S}^{-1} (\mathbf{x} - \mathbf{w}). \quad (\text{J.27})$$

7.9 Harmonic Mean

When related to example data and PVs, the Harmonic Mean measure in (I.12) becomes

$$d_{HM} = \sum_{i=1}^{N_F} \frac{x_i w_i}{x_i + w_i} \quad (\text{J.28})$$

on which one can use the quotient rule in (E.6) to compute the derivative with $u = x_i w_i$, $v = x_i + w_i$, and then for the derivative with respect to w : $du = x_i$ and $dv = 1$.

Therefore the derivative via the quotient rule is,

$$\frac{\partial d_{HM}}{\partial w} = \sum_{i=1}^{N_F} \frac{x_i(x_i + w_i) - x_i w_i}{(x_i + w_i)^2} \quad (\text{J.29})$$

APPENDIX K: Design of Experiments Results

Count what is countable, measure what is measurable, and what is not measureable, make measurable.

–GALILEO GALILEA, 1564 – 1642

Table K-1 presents design of experiments results for the cosine GRLVQI, Canberra GRLVQI, and Squared Euclidean GRLVQI (baseline) when considering all design points from Table V-6 for Z-Wave data. In Table K-1, factor levels correspond to those listed in Table V-6 with the notation of “–” for a low setting, “+” for a high setting, and “0” for the middle setting.

Table K-1: Design of Experiments Results

FACTOR					ALGORITHM								
					COSINE			CANBERRA			SQUARED EUCLIDEAN		
A	B	C	D	E	TRAIN AUCC	TEST AUCC	MEAN AUTH. AUC	TRAIN AUCC	TEST AUCC	MEAN AUTH. AUC	TRAIN AUCC	TEST AUCC	MEAN AUTH. AUC
-	-	-	-	-	13.22029	13.2029	0.974386	8.788406	7.846377	0.476263	14.68116	14.84203	0.736326
-	-	-	-	+	13.20725	13.22174	0.96775	8.773913	8.068116	0.572325	14.79565	14.74638	0.713711
-	-	-	0	-	13.49275	12.99565	0.987486	8.763768	8.001449	0.580113	14.65797	14.87681	0.740485
-	-	-	0	+	13.35797	13.2913	0.968299	8.8	8.042029	0.53436	14.68986	14.77681	0.690756
-	-	-	+	-	13.23623	13.28986	0.972098	8.795652	7.844928	0.546301	14.61884	14.94058	0.695009
-	-	-	+	+	13.28116	13.19565	0.966144	8.557971	7.981159	0.553403	14.64783	14.76957	0.688217
-	-	0	-	-	13.42029	13.36667	0.96017	8.775362	8.078261	0.513428	14.77101	14.78406	0.686377
-	-	0	-	+	13.31884	13.22174	0.970473	8.724638	8.004348	0.566093	14.64348	14.75072	0.693384
-	-	0	0	-	13.29565	13.41884	0.975728	8.763768	8.06087	0.579855	14.63333	14.85942	0.693422
-	-	0	0	+	13.33188	13.12319	0.934934	8.844928	8.1	0.628444	14.23188	14.47681	0.658381
-	-	0	+	-	13.3029	13.33043	0.990454	8.569565	8.002899	0.4754	14.23623	14.6058	0.855892
-	-	0	+	+	13.22174	13.17391	0.959168	8.708696	8.197101	0.514171	14.27536	14.49855	0.788733
-	-	+	-	-	13.23913	13.15797	0.954915	8.673913	8.163768	0.543371	14.30145	14.51304	0.838362
-	-	+	-	+	13.44638	13.57391	0.948204	8.913043	8.308696	0.518922	14.22899	14.21884	0.839609

-	-	+	0	-	13.42609	13.07391	0.971569	8.785507	7.995652	0.529975	14.23043	14.4913	0.801632
-	-	+	0	+	13.48261	13.12609	0.979244	8.74058	8.03913	0.589099	14.22899	14.36087	0.861934
-	-	+	+	-	13.43623	13.34203	0.954783	8.733333	7.913043	0.574304	14.33188	14.43768	0.807681
-	-	+	+	+	13.39565	13.52319	0.827202	8.723188	7.975362	0.636213	14.21014	14.08696	0.860328
-	0	-	-	-	13.36087	13.33333	0.972873	8.804348	8.086957	0.3739	14.21884	14.58841	0.84683
-	0	-	-	+	13.4029	13.48551	0.957656	8.657971	7.982609	0.523837	14.20145	14.43913	0.782086
-	0	-	0	-	13.17101	13.33768	0.981531	8.846377	7.868116	0.469074	14.28551	14.4058	0.829666
-	0	-	0	+	13.39855	13.29565	0.939376	8.768116	8.052174	0.646112	14.41014	14.26812	0.582237
-	0	-	+	-	13.61159	13.35072	0.952533	8.84058	8.133333	0.624411	14.34493	14.42899	0.675041
-	0	-	+	+	13.32174	13.33913	0.981456	8.77971	8.144928	0.444619	14.31304	14.37246	0.583264
-	0	0	-	-	13.50725	13.4	0.985501	8.788406	8.042029	0.500252	14.31014	14.41739	0.784871
-	0	0	-	+	13.55507	13.24058	0.989395	8.842029	7.818841	0.522602	14.3058	14.44928	0.832911
-	0	0	0	-	13.50435	13.31014	0.991607	8.665217	8.057971	0.59109	14.32319	14.34493	0.713163
-	0	0	0	+	15.16377	15.14928	0.993403	8.768116	7.823188	0.487561	15.12319	15.09565	0.912955
-	0	0	+	-	15.34203	15.05942	0.993062	8.673913	7.797101	0.65816	14.9971	15.15507	0.899477
-	0	0	+	+	15.28696	15.05072	0.995331	8.595652	7.898551	0.639729	14.87971	15.13043	0.89339
-	0	+	-	-	15.33043	15.17101	0.99448	8.84058	7.805797	0.695866	15.62029	15.54058	0.952949
-	0	+	-	+	15.15072	15.28116	0.995123	8.765217	7.991304	0.587694	15.62899	15.46232	0.947366
-	0	+	0	-	15.37391	15.28551	0.99518	8.650725	7.921739	0.719408	15.54348	15.52754	0.950208
-	0	+	0	+	15.29275	15.38116	0.994972	8.77971	7.904348	0.539698	15.35797	15.37681	0.957631
-	0	+	+	-	15.23043	15.27101	0.994764	8.818841	7.946377	0.591651	15.54493	15.56232	0.958916
-	0	+	+	+	15.30725	15.3058	0.99482	8.781159	7.984058	0.496988	15.57391	15.40725	0.960025
-	+	-	-	-	15.37681	15.2942	0.994972	8.856522	8.030435	0.438368	14.56957	14.56232	0.933749
-	+	-	-	+	15.24058	15.20725	0.993535	8.97971	7.947826	0.701393	14.78116	14.48696	0.912634
-	+	-	0	-	15.31014	15.14348	0.995066	9.081159	7.865217	0.75753	14.42029	14.39855	0.505898
-	+	-	0	+	15.35652	15.21884	0.995369	8.855072	7.943478	0.664858	15.02464	15.24928	0.94603
-	+	-	+	-	15.34928	15.24783	0.993459	8.756522	7.872464	0.685469	15.06377	15.15217	0.934121
-	+	-	+	+	15.19565	15.1087	0.993705	8.691304	8.062319	0.641613	15.05797	14.96087	0.939824
-	+	0	-	-	15.35217	15.32464	0.993573	8.981159	7.84058	0.53063	15.16087	15.07681	0.954631
-	+	0	-	+	15.20145	15.06812	0.995028	8.681159	8.266667	0.642974	15.02754	15.2913	0.94758
-	+	0	0	-	15.30145	15.31884	0.994272	8.815942	7.734783	0.65189	15.05797	14.9971	0.501985
-	+	0	0	+	15.41739	15.33478	0.994083	9.050725	8.16087	0.562187	15.11014	15.16812	0.923856
-	+	0	+	-	15.26667	15.42319	0.993724	8.886957	8.130435	0.545652	15.15072	15.24928	0.513882
-	+	0	+	+	15.23478	15.07681	0.992174	8.637681	7.992754	0.536749	14.94638	14.68841	0.517171
-	+	+	-	-	15.32029	15.33333	0.992401	8.911594	7.917391	0.690876	15.42174	15.43478	0.751449

-	+	+	-	+	15.33623	15.1942	0.995444	8.691304	8.033333	0.800567	15.34638	15.50435	0.512533
-	+	+	0	-	15.2942	15.27826	0.994197	8.989855	8.134783	0.625911	15.35797	15.37536	0.951323
-	+	+	0	+	15.27391	14.98116	0.995161	8.724638	8.095652	0.758652	15.36957	15.27391	0.949011
-	+	+	+	-	15.14783	15.09275	0.995406	8.84058	7.968116	0.443396	15.35652	15.36957	0.51167
-	+	+	+	+	15.33768	15.21159	0.995104	8.705797	7.986957	0.787832	15.24928	15.09565	0.775098
0	-	-	-	-	15.66522	15.47681	0.994008	8.653623	7.931884	0.554909	15.68261	14.96522	0.526415
0	-	-	-	+	15.69855	15.46087	0.994291	8.76087	7.753623	0.885192	15.67681	15.31594	0.971594
0	-	-	0	-	15.47536	15.62029	0.993686	8.531884	7.83913	0.745419	15.5	15.05217	0.968538
0	-	-	0	+	15.66377	15.76667	0.994272	8.592754	7.972464	0.545236	15.90435	15.34348	0.972111
0	-	-	+	-	15.62174	15.52319	0.99172	8.591304	8.097101	0.55666	15.9058	15.48841	0.521355
0	-	-	+	+	15.55652	15.48116	0.994159	8.626087	7.749275	0.848261	15.68261	15.39275	0.521588
0	-	0	-	-	15.52754	15.61739	0.992779	8.831884	7.933333	0.871374	15.92464	15.66812	0.522035
0	-	0	-	+	15.41594	15.28551	0.993705	8.730435	8.134783	0.583945	16.0971	15.89565	0.976434
0	-	0	0	-	15.65797	15.48986	0.993497	8.67971	7.989855	0.700076	15.96377	15.75652	0.968796
0	-	0	0	+	15.45072	15.0971	0.994026	8.62029	7.950725	0.804096	15.75072	15.22174	0.950788
0	-	0	+	-	15.53478	15.84928	0.993913	8.56087	8.095652	0.844587	15.65507	15.15942	0.970567
0	-	0	+	+	15.50725	15.43623	0.993138	8.94058	7.869565	0.80586	15.7	15.03188	0.969225
0	-	+	-	-	15.58841	15.8087	0.994291	8.65942	8.030435	0.6231	16.02174	15.54493	0.97172
0	-	+	-	+	15.51159	15.52464	0.992987	8.572464	8.13913	0.608532	16.08261	15.56812	0.955564
0	-	+	0	-	15.66522	15.48696	0.994802	8.782609	8.1	0.653856	15.97826	15.46812	0.647706
0	-	+	0	+	15.49565	15.32464	0.994612	8.830435	8.295652	0.538929	16.14348	15.7	0.640888
0	-	+	+	-	15.53768	15.4971	0.994858	8.57971	8.111594	0.570164	16.21594	15.91304	0.977624
0	-	+	+	+	15.48261	15.43913	0.990624	8.437681	7.905797	0.833245	16.08406	15.81739	0.977045
0	0	-	-	-	15.59275	15.6913	0.994594	8.968116	8.068116	0.902936	15.69565	15.53478	0.964436
0	0	-	-	+	15.46957	15.46957	0.994594	9.050725	7.934783	0.643648	15.82029	15.33913	0.958771
0	0	-	0	-	15.70435	15.66957	0.992987	8.744928	8.37971	0.819786	15.73043	15.47971	0.963207
0	0	-	0	+	15.52029	15.52464	0.989168	8.952174	8.157971	0.921739	16.16087	15.8942	0.965848
0	0	-	+	-	15.52899	15.30145	0.98913	8.768116	8.436232	0.780662	16.11304	15.66667	0.97925
0	0	-	+	+	15.54058	15.68841	0.995085	8.769565	8.107246	0.795885	16.21739	15.73043	0.979382
0	0	0	-	-	15.45507	15.33333	0.993289	8.810145	8.063768	0.673611	16.2971	16.1	0.966194
0	0	0	-	+	15.52174	15.22899	0.993648	8.762319	7.944928	0.804392	16.30145	15.75507	0.974026
0	0	0	0	-	15.65652	15.24928	0.992892	8.824638	8.05942	0.767914	16.1058	15.9	0.823485
0	0	0	0	+	13.13478	13.37101	0.980132	8.721739	8.173913	0.448727	14.53188	14.04638	0.770592
0	0	0	+	-	13.16957	12.76667	0.911682	8.617391	8.163768	0.572344	14.54638	14.15217	0.775381
0	0	0	+	+	13.04058	13.05942	0.981947	8.881159	8.011594	0.608551	14.44058	13.87681	0.785784

0	0	+	-	-	13.03768	12.91884	0.973837	8.807246	8.088406	0.595167	14.5	13.9029	0.648815
0	0	+	-	+	13.11159	12.97971	0.972136	8.721739	7.642029	0.521222	14.62174	14.38551	0.621109
0	0	+	0	-	12.94783	13.21304	0.954726	8.686957	8.176812	0.516982	14.56812	14.1087	0.677032
0	0	+	0	+	13.17971	13.01884	0.983913	8.747826	8.007246	0.458147	14.61304	14.25072	0.641078
0	0	+	+	-	12.83623	12.9971	0.98104	8.749275	8.014493	0.58402	14.54493	14.31014	0.643869
0	0	+	+	+	13.10725	13.07391	0.921626	8.746377	8.046377	0.45804	14.55507	14.31449	0.62414
0	+	-	-	-	12.87536	13.04493	0.897826	8.728986	8.244928	0.475079	14.23768	14.12899	0.830662
0	+	-	-	+	12.98551	13.18116	0.890095	8.869565	8.165217	0.557001	14.23623	14.01594	0.789137
0	+	-	0	-	13.04783	13.19855	0.984839	8.730435	7.905797	0.493239	14.35217	13.63768	0.835854
0	+	-	0	+	13.02174	13.26667	0.985992	8.723188	8.133333	0.578544	14.2087	13.85652	0.767832
0	+	-	+	-	12.99855	13.05362	0.985142	9.023188	7.963768	0.438185	14.28696	13.96812	0.760038
0	+	-	+	+	12.9913	13.25072	0.965369	8.733333	8.126087	0.499408	14.33333	13.91594	0.783793
0	+	0	-	-	13.0913	13.07971	0.98673	8.692754	7.798551	0.454726	13.96377	13.81884	0.75644
0	+	0	-	+	12.99855	13.27391	0.850473	8.72029	8.124638	0.544921	14.15217	13.84058	0.739382
0	+	0	0	-	12.98986	13.01884	0.678544	8.853623	7.988406	0.487026	14.3058	13.90145	0.740964
0	+	0	0	+	12.98261	12.80145	0.972042	8.615942	8.110145	0.67828	14.38696	13.84493	0.746553
0	+	0	+	-	12.77101	13.04493	0.966616	8.588406	8.178261	0.511424	14.37101	14.04638	0.769962
0	+	0	+	+	13.10435	12.87246	0.951096	8.824638	7.92029	0.71988	14.35362	13.76232	0.802684
0	+	+	-	-	12.95362	12.98116	0.951664	8.765217	7.989855	0.498009	14.3971	14.08406	0.680725
0	+	+	-	+	12.93043	13.0942	0.97155	8.608696	7.9	0.533056	14.36522	14.13043	0.751474
0	+	+	0	-	13.12319	12.93188	0.898431	8.757971	8.047826	0.508513	14.43333	14.00145	0.768425
0	+	+	0	+	12.93768	13.30435	0.892363	8.689855	7.837681	0.529817	14.26087	13.86667	0.727839
0	+	+	+	-	13.06522	12.78261	0.947732	8.766667	7.985507	0.543314	14.42899	13.99275	0.731235
0	+	+	+	+	12.8942	12.97246	0.957372	8.694203	8.028986	0.470082	14.46957	14.22319	0.651159
+	-	-	-	-	15.05217	14.99275	0.989943	8.724638	7.788406	0.525829	14.91884	15.01304	0.9177
+	-	-	-	+	15.24348	15.3087	0.993403	8.82029	8.068116	0.511487	14.87391	14.63623	0.91264
+	-	-	0	-	15.26667	15.12319	0.994896	8.818841	8.107246	0.449112	14.67391	14.38696	0.905325
+	-	-	0	+	15.13623	14.90145	0.993535	8.795652	8.036232	0.78564	15.21884	14.88261	0.955734
+	-	-	+	-	15.2913	15.01159	0.989603	8.714493	8.02029	0.570744	15.35362	15.3087	0.946049
+	-	-	+	+	15.21014	15.05652	0.995595	8.649275	8.014493	0.793377	14.97101	14.67246	0.933951
+	-	0	-	-	15.02754	14.88986	0.994915	8.723188	8.024638	0.677013	15.17971	15.07391	0.947467
+	-	0	-	+	15.1913	15.12899	0.995482	8.630435	7.72029	0.722111	15.23623	15.11594	0.953214
+	-	0	0	-	15.12609	15.44203	0.995652	8.844928	7.982609	0.438129	15.29275	15.31449	0.933106
+	-	0	0	+	15.1058	15.22899	0.993667	8.86087	7.913043	0.669666	14.36087	14.2	0.905608
+	-	0	+	-	15.20145	15.26522	0.994915	8.715942	8.075362	0.538834	14.4058	14.17536	0.923113

+	-	0	+	+	15.11884	15.0971	0.995652	8.647826	7.942029	0.707763	14.17536	13.76522	0.862029
+	-	+	-	-	15.2087	15.19855	0.995652	8.801449	8.108696	0.698355	14.96232	14.75217	0.911229
+	-	+	-	+	15.20435	15.15797	0.994858	8.882609	8.044928	0.597379	14.85507	14.5971	0.93765
+	-	+	0	-	15.24493	15.23188	0.995388	8.798551	8.026087	0.720252	14.56087	14.33188	0.891487
+	-	+	0	+	15.23623	15.22174	0.995652	8.889855	7.965217	0.535696	15.07536	14.38841	0.514852
+	-	+	+	-	15.24783	15.31884	0.994159	8.884058	7.844928	0.796497	14.86522	14.77826	0.95528
+	-	+	+	+	15.09855	15.01014	0.99552	9.06087	8.047826	0.534127	14.70435	14.45072	0.944644
+	0	-	-	-	15.25797	15.0971	0.994972	8.833333	7.823188	0.455041	14.95652	14.72754	0.912483
+	0	-	-	+	15.17826	15.12609	0.994405	8.881159	7.965217	0.546541	14.56377	14.31884	0.852848
+	0	-	0	-	15.23478	14.98696	0.995369	8.672464	7.943478	0.613611	14.56667	14.16522	0.891771
+	0	-	0	+	15.41304	15.20725	0.995085	8.775362	8.101449	0.486761	15.33188	15.21884	0.51443
+	0	-	+	-	15.25217	15.18116	0.994026	8.675362	8.131884	0.686957	15.36667	15.28696	0.520586
+	0	-	+	+	15.20435	15.14058	0.992552	8.911594	7.910145	0.687618	14.93478	14.68261	0.923409
+	0	0	-	-	15.41014	15.12174	0.995028	8.75942	8.114493	0.523428	15.33913	15.10435	0.945904
+	0	0	-	+	15.2942	15.24638	0.994442	8.689855	8.121739	0.539351	15.31159	15.39275	0.955986
+	0	0	0	-	15.21014	15.27246	0.995406	8.876812	7.986957	0.68138	15.29565	14.77536	0.912703
+	0	0	0	+	15.51884	15.51014	0.994216	8.878261	7.981159	0.611853	15.23043	14.75217	0.522741
+	0	0	+	-	15.49855	15.54638	0.992533	8.865217	7.975362	0.5177	15.34058	14.75217	0.517297
+	0	0	+	+	15.67826	15.55652	0.994253	8.465217	7.913043	0.615482	15.26812	14.3942	0.968053
+	0	+	-	-	15.55652	15.79855	0.993081	8.531884	7.884058	0.796043	15.85362	15.21739	0.974348
+	0	+	-	+	15.53623	15.6029	0.993403	8.662319	8.055072	0.513711	15.58696	14.82899	0.517133
+	0	+	0	-	15.47391	15.45797	0.993913	8.721739	8.005797	0.585167	15.58261	14.69565	0.974008
+	0	+	0	+	15.41159	15.4942	0.993875	8.775362	7.866667	0.919086	15.65072	15.31594	0.974858
+	0	+	+	-	15.46957	15.62319	0.984008	8.850725	7.965217	0.549572	15.66957	15.33623	0.517694
+	0	+	+	+	15.59275	15.31014	0.994631	8.821739	7.95942	0.723686	15.5058	15.02029	0.976736
+	+	-	-	-	15.51159	15.58406	0.993176	8.821739	8.049275	0.610315	15.72609	15.09855	0.961582
+	+	-	-	+	15.67246	15.55072	0.994178	8.791304	7.843478	0.765142	15.69275	14.69855	0.607057
+	+	-	0	-	15.42029	15.55652	0.990548	8.75942	8.104348	0.777372	15.38116	14.69855	0.639326
+	+	-	0	+	15.46812	15.47826	0.99482	8.682609	8.001449	0.626068	15.66232	14.87971	0.964354
+	+	-	+	-	15.47826	15.32174	0.993648	8.765217	7.911594	0.842054	15.61739	14.96957	0.592035
+	+	-	+	+	15.61159	15.56812	0.994348	8.813043	8.073913	0.561103	15.58986	14.68261	0.596673
+	+	0	-	-	15.62319	15.47536	0.993648	8.763768	8.273913	0.615406	15.67826	15.39565	0.937883
+	+	0	-	+	15.64638	15.44058	0.993667	8.627536	8.098551	0.576522	15.86377	15.52899	0.594127
+	+	0	0	-	15.68116	15.49275	0.99327	8.723188	7.984058	0.757328	15.65072	15.16377	0.9715
+	+	0	0	+	15.62029	15.23478	0.99552	8.736232	8.291304	0.600989	15.45652	14.84783	0.852489

+	+	0	+	-	15.39855	15.61884	0.993554	8.830435	8.118841	0.767883	15.37681	15.04203	0.75903
+	+	0	+	+	15.65217	15.1942	0.993006	8.836232	8.310145	0.718973	15.35942	14.90435	0.969975
+	+	+	-	-	15.47971	15.16232	0.978431	8.737681	8.127536	0.77804	16.01159	15.42319	0.970838
+	+	+	-	+	15.51449	15.41449	0.986522	8.814493	8.256522	0.826307	15.7087	15.44928	0.969036
+	+	+	0	-	15.54058	15.40725	0.989773	8.727536	7.963768	0.790945	15.74928	15.02899	0.949389
+	+	+	0	+	15.48261	15.5087	0.992779	8.791304	7.87971	0.559748	15.98986	15.64493	0.958551
+	+	+	+	-	15.57391	15.44783	0.981399	8.589855	8.228986	0.708135	16.06667	15.71739	0.644052
+	+	+	+	+	15.56087	15.46087	0.994499	8.927536	8.104348	0.877587	15.74928	15.37826	0.975009

APPENDIX L: MEX File Programming Considerations

A computer is like an Old Testament god, with a lot of rules and no mercy.

–JOSEPH CAMPBELL, 1904 – 1987

For efficiency, MATLAB mex files are used for GRLVQI implementation on RF-DNA data. Writing mex files involves understanding both Matlab and C programming. Common programming issues encountered with mex files included: 1) improper distinctions between pointers and variables in the mex file, 2) complexities and differences in mathematical programming that exist between Matlab and C.

Additionally, compiling mex files appropriately is nontrivial. While the below syntax will compile a mex file, not all mex files performed equally fast and hence the computational speed of a mex file appears to have a connection to the computer and software it was compiled on. Per communication with Reising [599], for debugging and coding considerations one should compile a given mex file via the following commands:

$$\begin{aligned} \text{mex -g -v COMPFLAGS} \\ = "\$COMPFLAGS -Wall" - \text{largeArrayDims FILENAME.c} \end{aligned} \quad (\text{H.1})$$

where compiling with the “-g” command enables debugging in Microsoft Visual Studio [600].

For debugging a given mex file one should consider the following general process:

1. Start Matlab
2. Compile

3. Start Microsoft Visual Studio
4. Open the associated c-file in Microsoft Visual Studio
5. Attach Microsoft Visual Studio to the Matlab process
6. Insert break points as needed in the c file (within Visual Studio)
7. Run the Matlab algorithm under analysis.

When these steps are followed, one will find that Matlab and Visual Studio enable rough debugging abilities of mex files.

APPENDIX M: GRLVQI-D Performance on ZigBee RF-DNA Fingerprints with Z-Wave Based Optimization

Beware that thou be not deceived into folly, and be humbled.

–SIRACH 13:10 (DRA)

ZigBee data was also considered using the optimized Squared Euclidean GRLVQI and the optimized Cosine GRLVQI-D algorithms. However, it should be noted that the optimized settings are only optimized per Z-Wave RF-DNA fingerprints and thus no guarantees on their applicability to ZigBee. Future research item number 2, in Section 7.3, regards using the Air Force Research Laboratory DOD Supercomputing Resource Center (DSRC). This is directly connected to the results in this appendix. Due to computational times associated with the larger ZigBee dataset (when compared to the Z-Wave dataset), the optimization process was not reconsidered for ZigBee devices. Additionally, since the Canberra GRLVQI algorithmic results generally underperformed both the Squared Euclidean GRLVQI and Cosine GRLVQI-D, Canberra GRLVQI-D was not further considered for ZigBee RF-DNA Fingerprints.

Figure M-1 presents training (TNG) and testing (TST) classification results from the baseline Squared Euclidean GRLVQI algorithm, the Squared Euclidean GRLVQI algorithm using the Classification-optimized settings in Table V-9, and the Squared Euclidean GRLVQI algorithm using the Verification-optimized settings in Table V-9. Noticeably, classification performance of the optimized algorithms appears slightly lower than the baseline ZigBee GRLVQI performance. The Classification-based optimized

Squared Euclidean GRLVQI shows an improvement in gain of -4.4 dB (TNG) and -2.69 dB (TST) at 90% accuracy; the Verification-based optimized Squared Euclidean GRLVQI shows an improvement in gain of -13.44 dB (TST) and -10.48 dB (TST).

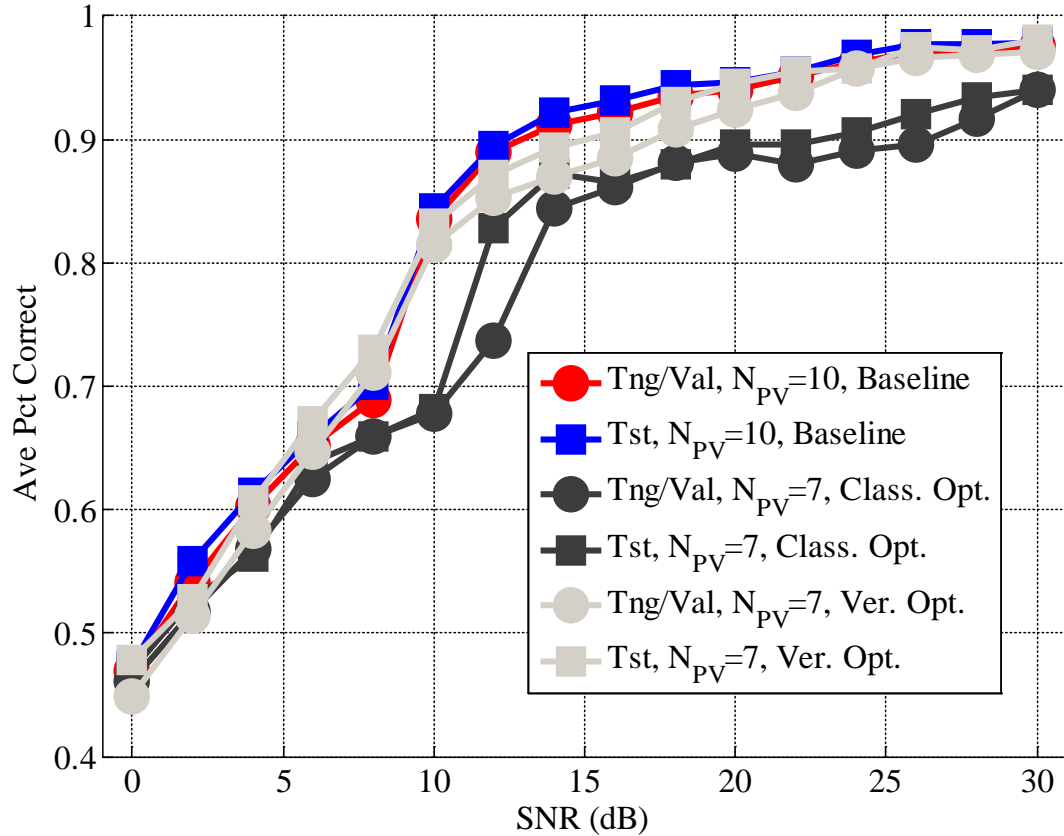


Figure M-1: ZigBee GRLVQI Classification Performance Using Squared Euclidean Distance Using Optimized Algorithmic Settings.

Figure M-2 presents both the authorized, Figure M-2a, and rogue rejected, Figure M-2b, verification performance for the Classification-optimized Squared Euclidean GRLVQI algorithm. When compared with baseline performance, presented in Table V-5, the Classification-optimized Squared Euclidean GRLVQI performance has

improved authorized verification performance (50% versus 25%), but reduced rogue rejection verification performance (30.56% versus 52.78%).

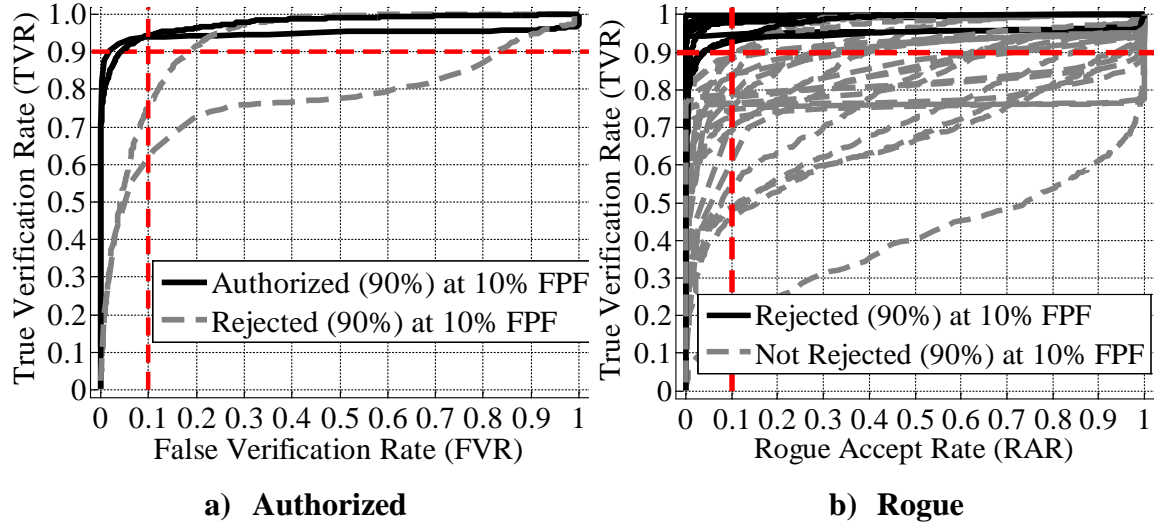


Figure M-2: GRLVQI ID Verification Performance of ZigBee in Squared Euclidean GRLVQI using Z-Wave Determined Classification-Based Optimization Settings at 18dB.

Figure M-3 similarly presents both the authorized, Figure M-3a, and rogue rejected, Figure M-3b, verification performance for the Verification-optimized Squared Euclidean GRLVQI algorithm. Noticeably, performance is degraded compared to the Classification-optimized algorithmic results in Figure M-2. When compared with baseline performance, presented in Table V-5, the Classification-optimized Squared Euclidean GRLVQI performance has worse authorized verification performance (0% versus 25%), and worse rogue rejection verification performance (41.66% versus 52.78%).

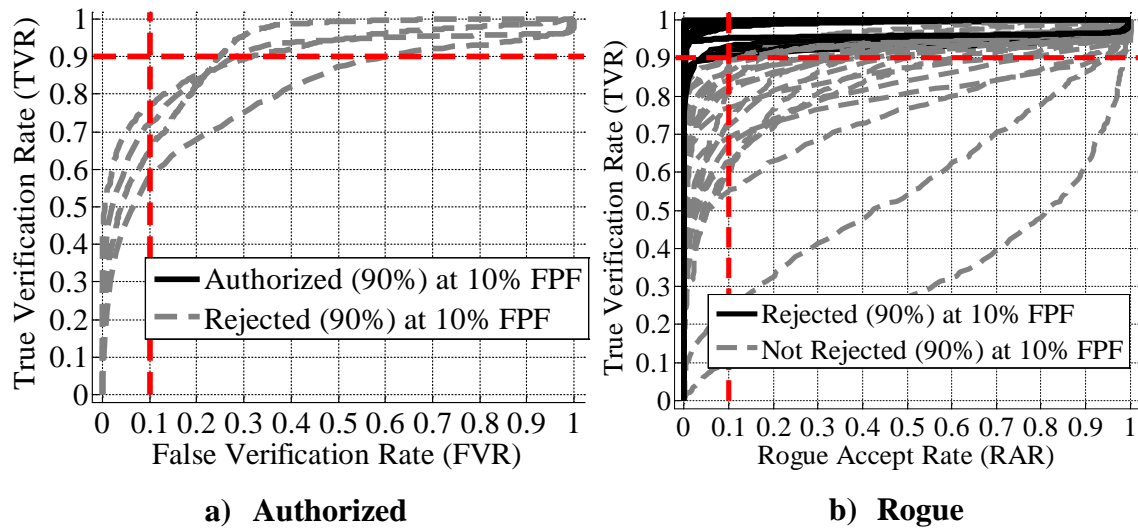


Figure M-3: GRLVQI ID Verification Performance of ZigBee in Squared Euclidean GRLVQI using Z-Wave Determined Vefication-Based Optimization Settings at 18dB.

Figure M-4 presents training (TNG) and testing (TST) classification results from the Cosine GRLVQI-D algorithm in comparison with the baseline Squared Euclidean GRLVQI algorithm. Both Cosine GRLVQI-D with the Classification-optimized settings in Table V-9 and the Cosine GRLVQI algorithm using the Verification-optimized settings in Table V-9 are presented. Noticeably, classification performance of the optimized algorithms appears slightly worse than the baseline ZigBee GRLVQI performance and performance never reaches 90% accuracy.

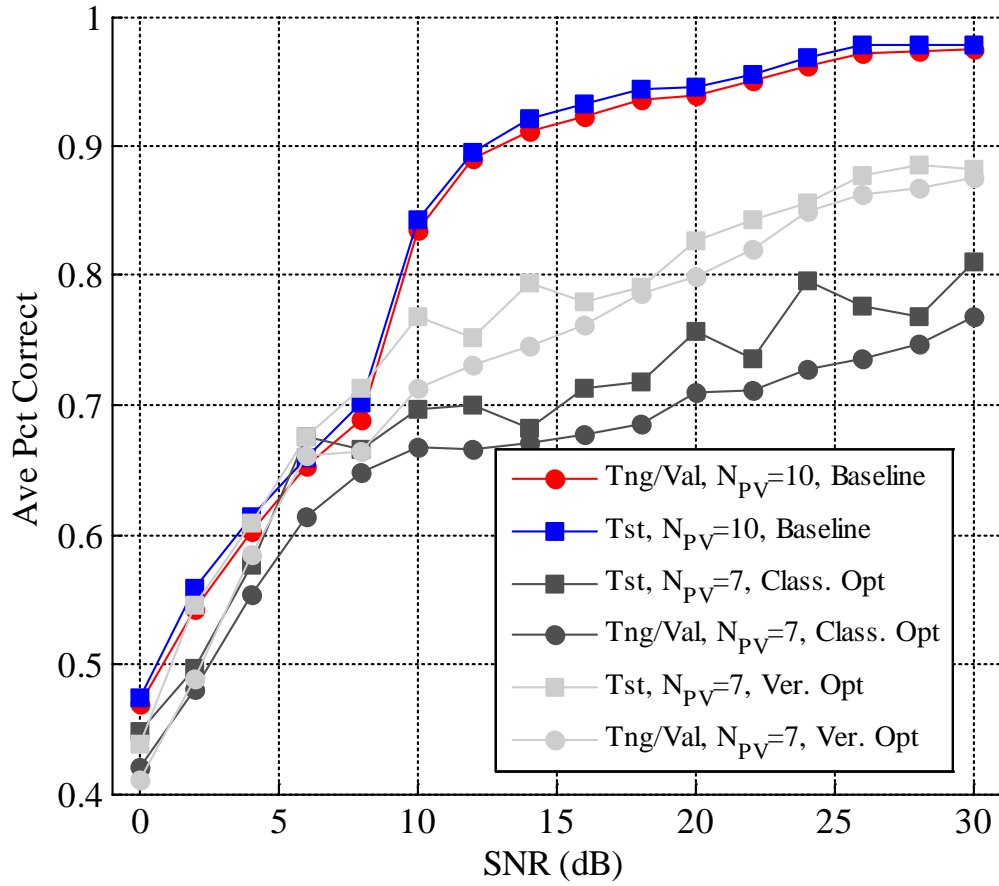


Figure M-4: GRLVQI Classification Performance Using Cosine Distance Using Optimized Algorithmic Settings.

Figure M-5 presents both the authorized, Figure M-5a, and rogue rejected, Figure M-5b, verification performance for the Classification-optimized Cosine GRLVQI-D algorithm. When compared with baseline Squared Euclidean GRLVQI performance, presented in Table V-5, the Classification-optimized Cosine GRLVQI-D performance has comparable authorized verification performance (25% versus 25%), but reduced rogue rejection verification performance (47.22% versus 52.78%).

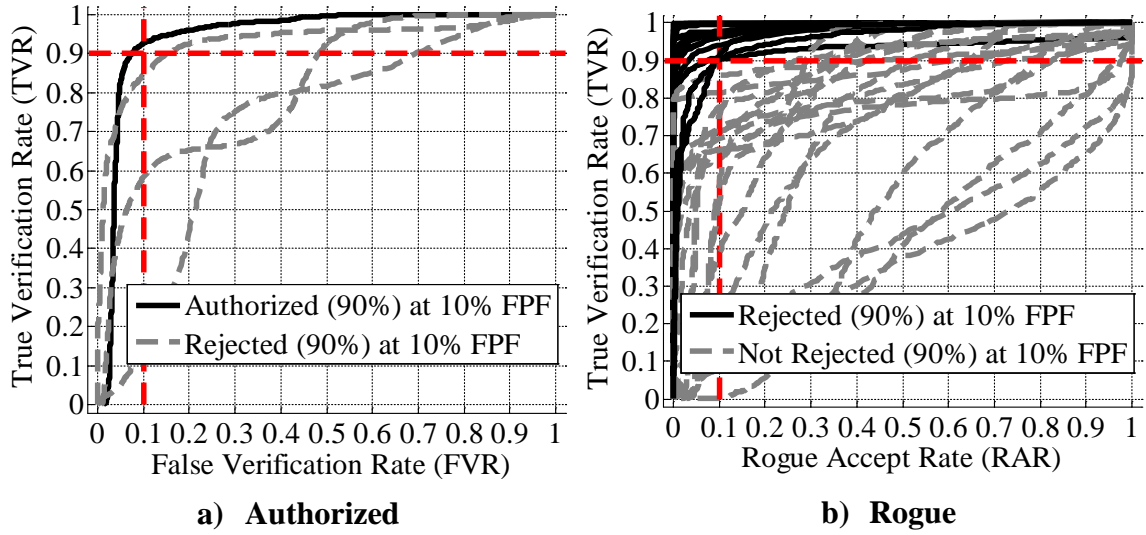


Figure M-5: GRLVQI ID Verification Performance of ZigBee in Cosine GRLVQI using Z-Wave Determined Classification-Based Optimization Settings at 18dB.

Figure M-6 similarly presents both the authorized, Figure M-6a, and rogue rejected, Figure M-6b, verification performance for the Verification-optimized Cosine GRLVQI-D algorithm. Noticeably, performance is slightly degraded compared to the Classification-optimized algorithmic results in Figure M-5, which is consistent with the observations about Squared Euclidean GRLVQI in Figure M-2 and Figure M-3 . When compared with baseline performance, presented in Table V-5, the Classification-optimized Squared Euclidean GRLVQI performance has worse authorized verification performance (0% versus 25%), and worse rogue rejection verification performance (33.33% versus 52.78%).

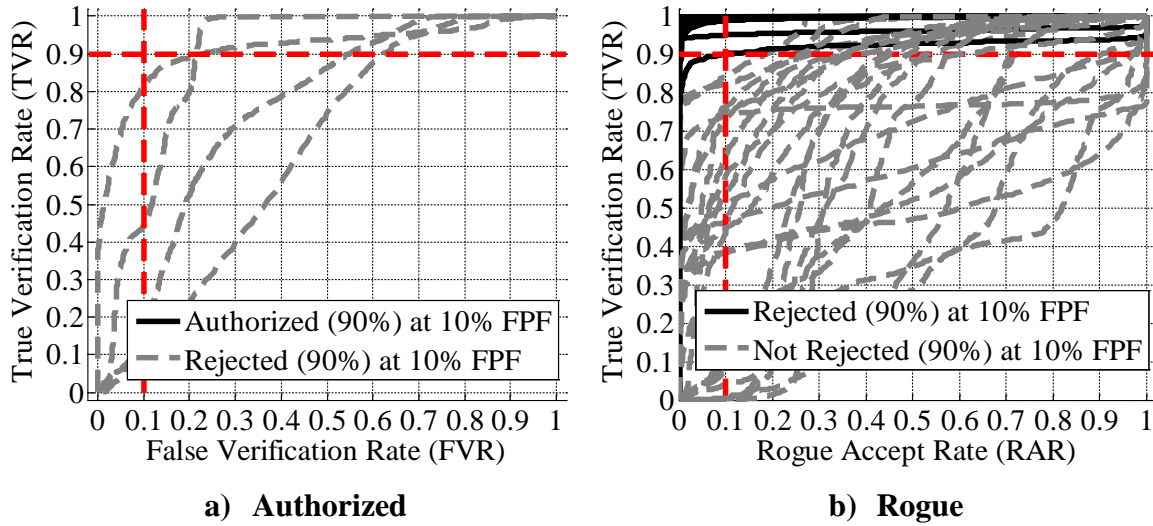


Figure M-6: GRLVQI ID Verification Performance of ZigBee in Cosine GRLVQI using Z-Wave Determined Verification-Based Optimization Settings at 18dB.

Table M-1 presents an overall comparison of classification and verification performance for the Squared Euclidean GRLVQI algorithm and the Cosine GRLVQI-D algorithm. Baseline performance from Table V-5 is also included for comparison. Overall, the best performance is seen in the non-optimized Squared Euclidean GRLVQI algorithms. This differs from the result seen in Section 5.4.3 when the Z-Wave dataset was considered.

Table M-1: GRLVQI Performance for ZigBee RF-DNA Data Using Z-Wave Optimized Algorithmic Settings.

RESULT									
ALGORITHM	OPTIMIZATION METHOD	CLASSIFICATION				VERIFICATION (18 dB)			
		RAP (TNG)	RAP (TST)	SNR GAIN (dB) AT 90% C RELATIVE TO BASELINE TST ($N_{PV}=10$)		% AUTHORIZE D OR % ROGUE REJECTED		MEAN AUC	
SQUARED EUCLIDEAN GRLVQI	None – Baseline Settings ($N_{PV} = 10$)	0.99	1.00	−0.53	0.00	25%	63.9%	0.92	0.93
	None – Baseline Settings ($N_{PV} = 13$)	1.00	1.01	−0.11	+0.5	25%	52.8%	0.93	0.94
	Classification- Based Optimization	0.91	0.93	−4.93	−2.7	50%	30.6%	0.91	0.87
	Verification- Based Optimization	0.97	0.99	−13.9	−10.5	0%	41.7%	0.88	0.90
COSINE GRLVQI-D	Classification- Based Optimization	0.78	0.82	N/A		25%	47.2%	0.85	0.85
	Verification- Based Optimization	0.87	0.90	N/A		0%	33.3%	0.80	0.81

VIII. Bibliography

- [1] A. P. Snow, U. Varshney and A. D. Malloy, "Reliability and survivability of wireless and mobile networks," *Computer*, vol. 33, no. 7, pp. 49-55, 2000.
- [2] R. J. Ellison, D. A. Fisher, R. C. Linger, H. F. Lipson and T. Longstaff, "Survivable network systems: An emerging discipline," Software Engineering Ins., Pittsburgh, PA, 1997.
- [3] E. A. Luijff and M. H. Klaver, "Protecting a nation's critical infrastructure: The first steps," *IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, pp. 1185-1190, 2004.
- [4] H. R. McMaster, Crack in the Foundation: Defense Transformation and the Underlying Assumption of Dominant Knowledge in Future War, Carlisle, PA: US Army War College, 2003.
- [5] Q. Liao, X. R. Luo, A. Gurung and W. Shi, "A holistic understanding of non-users' adoption of university campus wireless network: An empirical investigation," *Computers in Human Behavior*, vol. 48, pp. 220-229, 2015.
- [6] S. Prabhakar, S. Pankanti and A. K. Jain, "Biometric recognition: Security and privacy concerns," *IEEE Security and Privacy*, pp. 33-42, March/April 2003.
- [7] M. R. Schonberg, Defining the DOD Role in National Cybersecurity, Carlisle Barracks, PA: Army War College Center for Strategic Leadership, 2013.
- [8] R. R. Schell, Computer Security: The Achilles' Heel of the Electronic Air Force?, Monterey, CA: Naval Postgraduate School, 2013.
- [9] L. H. Jordan Jr, S. Metz, R. J. Howell, D. Connelly, J. Melody, J. Surdu, K. Brischke, D. Cranz, J. Gailliard and B. C. Proctor, US Army War College Key Strategic Issues List (KSIL) 2012-2013., Carlisle Barracks, PA: Army War College, 2012.
- [10] M. Rostami, F. Koushanfar and R. Karri, "A Primer on Hardware Security: Models, Methods, and Metrics," *Proceedings of the IEEE*, vol. 102, no. 8,

pp. 1283-1295, 2014.

- [11] J. Villasenor and M. Tehranipoor, "Chop-Shop Electronics: Clever Counter," *IEEE Spectrum*, vol. 50, no. 10, pp. 41-45, 2013.
- [12] F. Koushanfar, S. Fazzari, C. McCants, W. Bryson, M. Sale, P. Song and M. Potkonjak, "Can EDA combat the rise of electronic counterfeiting?," *Annual Design Automation Conference*, pp. 133-138, 2012.
- [13] Y. Zheng, X. Wang and S. Bhunia, "SACCI: Scan-Based Characterization Through Clock Phase Sweep for Counterfeit Chip Detection," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, pp. 1-11, 2014.
- [14] H. Huang, A. Boyer and S. B. Dhia, "The detection of counterfeit integrated circuit by the use of electromagnetic," *EMC Europe*, pp. 1-5, 2014.
- [15] K. Huang, J. M. Carulli and Y. Makris, "Counterfeit electronics: A rising threat in the semiconductor manufacturing industry," *IEEE International Test Conference*, pp. 1-4, 2013.
- [16] U. Guin, D. DiMase and M. Tehranipoor, "Counterfeit integrated circuits: Detection, avoidance, and the challenges ahead," *Journal of Electronic Testing*, vol. 30, no. 1, pp. 9-23, 2014.
- [17] A. Sreedhar, S. Kundu and I. Koren, "On Reliability Trojan Injection and Detection," *Journal of Low Power Electronics*, vol. 8, pp. 1-10, 2012.
- [18] W. E. Cobb, E. W. Garcia, M. A. Temple, R. O. Baldwin and Y. C. Kim, "Physical layer identification of embedded devices using RF-DNA fingerprinting," *Military Communications Conference (MILCOM)*, pp. 2168-2173, 2010.
- [19] W. E. Cobb, E. D. Laspe, R. O. Baldwin, M. A. Temple and Y. C. Kim, "Intrinsic physical-layer authentication of integrated circuits," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 1, pp. 14-24, 2012.
- [20] I. Montanari and A. Tacchini, "Aging Effects on Electromagnetic Signature of Electronic Devices," *IEEE International Symposium on Electromagnetic Compatibility and Electromagnetic Ecology*, pp. 159-161, 2007.
- [21] A. Boyer and S. B. Dhia, "Effect of Aging on Power Integrity and Conducted Emission of Digital Integrated Circuits," *Journal of Low Power Electronics*

, vol. 10, no. 1., pp. 165-172, 2014.

- [22] J. F. Dawson, I. D. Flintoft, A. P. Duffy, A. C. Marvin and M. P. Robinson, "Effect of high temperature ageing on electromagnetic emissions from a PIC microcontroller," *IEEE International Symposium on Electromagnetic Compatibility*, pp. 1139-1143, 2014.
- [23] A. P. Duffy, J. F. Dawson, I. D. Flintoft and A. C. Marvin, "Electromagnetic Monitoring of Semiconductor Ageing," *Procedia CIRP*, vol. 22, pp. 98-102, 2014.
- [24] A. Boyer and S. Ben Dhia, "Effect of electrical stresses on digital integrated circuits power integrity," *IEEE Workshop on Signal and Power Integrity (SPI)*, pp. 1-4, 2013.
- [25] A. Boyer and S. Ben Dhia, "Characterization and modeling of electrical stresses on digital integrated circuits power integrity and conducted emission.,", *International Workshop on Electromagnetic Compatibility of Integrated Circuits*, pp. 190-195, 2013.
- [26] I. Montanari, A. Tacchini and M. Maini, "Impact of thermal stress on the characteristics of conducted emissions," *IEEE International Symposium on Electromagnetic Compatibility*, pp. 1-4, 2008.
- [27] I. Montanari, "EMI measurements for aging control and fault diagnosis in active devices," *IEEE International Symposium on Electromagnetic Compatibility and Electromagnetic Ecology*, pp. 183-186, 2005.
- [28] H. J. Porck, Rate of paper degradation: the predictive value of artificial aging tests, Amsterdam: European Commission on Preservation and Access, 2000.
- [29] S. Zervos, "Natural and accelerated ageing of cellulose and paper: A literature review," *Cellulose: Structure and properties, derivatives and industrial uses*, pp. 155-203, 2010.
- [30] B. T. Hotle, J. M. Considine, M. J. Wald, R. E. Rowlands and K. T. Turner, "Effects of thermal aging on mechanical performance of paper," *In Finland: Progress in Paper Physics Seminar*, pp. 271-273, 2008.
- [31] F. X. Perrin, M. Irigoyen, E. Aragon and J. L. Vernet, "Evaluation of accelerated weathering tests for three paint systems: a comparative study of their aging behaviour," *Polymer degradation and stability*, vol. 72, no. 1, pp. 115-124,

2001.

- [32] M. Song and J. Munn, "Permanence, Durability and Unique Properties of Hanji," *Book and Paper Group Annual*, vol. 23, pp. 127-136, 2004.
- [33] J. R. Black, "Electromigration - A brief survey and some recent results," *IEEE Transactions on Electron Devices*, vol. 16, no. 4, pp. 338-347, 1969.
- [34] A. Kurokawa, M. Watanabe, M. Hoshi and M.-a. Fukase, "Reasonable circuit analysis considering comprehensively reliability and variability," *IEEE International Symposium on Communications and Information Technologies (ISCIT)*, pp. 282-287, 2013.
- [35] A. Grau, "Can you trust your fridge?," *IEEE Spectrum*, pp. 51-56, Mar. 2015.
- [36] A. Bari, J. Jiang, W. Saad and A. Jaekel, "Challenges in the Smart Grid Applications: An Overview," *International Journal of Distributed Sensor Networks*, vol. 2014, pp. 1-11, 2014.
- [37] J. Wright, "KillerBee: Practical zigbee exploitation framework," in *11th ToorCon Conf.*, San Diego, 2009.
- [38] J.-S. Lee, Y.-W. Su and C.-C. Shen, "A comparative study of wireless protocols: Bluetooth, UWB, ZigBee, and Wi-Fi," *33rd Annual Conference on the IEEE Industrial Electronics Society (IECON)*, pp. 46-51, 2007.
- [39] N. Baker, "ZigBee and Bluetooth strengths and weaknesses for industrial applications," *Comput. & Control Eng. J.*, vol. 16, no. 2, pp. 20-25, 2005.
- [40] J.-Y. Cheng, M.-H. Hung and J.-W. Chang, "A zigbee-based power monitoring system with direct load control capabilities," *IEEE International Conference on Networking, Sensing and Control*, pp. 895-900, 2007.
- [41] P. P. Parikh, M. G. Kanabar and T. S. Sidhu, "Opportunities and challenges of wireless communication technologies for smart grid applications," *IEEE Power and Energy Society General Meeting*, pp. 1-7, 2010.
- [42] K. S. Hung, W. K. Lee, V. O. K. Li, K. S. Lui, P. W. T. Pong, K. K. Y. Wong, G. H. Yang and J. Zhong, "On wireless sensors communication for overhead transmission line monitoring in power delivery systems," *IEEE International Conference on Smart Grid Communications*

(*SmartGridComm*), pp. 309-314, 2010.

- [43] Z. Zhao, W. C. Lee, Y. Shin and K.-B. Song, "An optimal power scheduling method for demand response in home energy management system," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1391-1400, 2013.
- [44] S.-K. Chen, T. Kao, C.-T. Chan, C.-N. Huang, C.-Y. Chiang, C.-Y. Lai, T.-H. Tung and P.-C. Wang, "A reliable transmission protocol for ZigBee-based wireless patient monitoring," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 1, pp. 6-16, 2012.
- [45] J. Hou, C. Wu, Z. Yuan, J. Tan, Q. Wang and Y. Zhou, "Research of intelligent home security surveillance system based on ZigBee," *Int. Symp. on Intelligent Inform. Technology Applicat. Workshops (IITAW)*, pp. 554-557, 2008.
- [46] D. Egan, "The emergence of ZigBee in building automation and industrial control," *Comput. & Control Eng. J.*, vol. 16, no. 2, pp. 14-19, 2005.
- [47] C. Reinisch, W. Kastner, G. Neugschwandtner and W. Granzer, "Wireless technologies in home and building automation," *IEEE International Conference on Industrial Informatics*, pp. 93-98, 2007.
- [48] V. Pandey, S. S. Gill and A. Sharma, "Wireless electricity theft detection system using ZigBee technology," *Int. J. Recent and Innovation Trends in Computing and Commun.*, vol. 1, no. 4, pp. 364-367, 2013.
- [49] T. J. Bihl, M. A. Temple, K. W. Bauer and B. Ramsey, "Dimensionality Reduction Analysis for Physical Layer Device Fingerprints with Application to ZigBee and Z-Wave Devices," *IEEE Military Communications Conference (MILCOM)*, pp. 360-365, 2015.
- [50] N. Edwards, Hardware intrusion detection for supply-chain threats to critical infrastructure embedded systems, University of Illinois at Urbana-Champaign: MS Thesis, 2013.
- [51] D. R. Reising, Exploitation of RF-DNA for device classification and verification using GRLVQI processing, Air Force Institute of Technology: PhD Dissertation, 2012.
- [52] Y. Chen, W. Trappe and R. Martin, "Detecting and localizing wireless spoofing attacks," *IEEE Conference on Sensor, Mesh and AdHoc Communications*

and Networks, pp. 193-202, 2007.

- [53] A. Leest and M. Bastiaans, "Gabor's discrete signal expansion and the discrete gabor transform on a non-separable lattice," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 101-104, 2000.
- [54] J. Neufeld, C. Fitfield, C. Doerr, A. Sheth and D. Grunwald, "SoftMAC: Flexible wireless research platform," *4th Workshop on Hot Topics in Networks*, 2005.
- [55] Y. Sheng, K. Tan, G. Chen, D. Kotz and A. Campbell, "Detecting 802.11 MAC layer spoofing using received signal strength," *27th Conference on Computer Communications*, 2008.
- [56] T. Tarman and E. Witzke, "Intrusion detection considerations for switched networks," *Proceedings of the SPIE: Enabling Technologies for Law Enforcement and Security*, vol. 4232, 2001.
- [57] D. Zhenhai, X. Yuan and J. Chandrashekar, "Controlling IP spoofing through interdomain packet filters," *IEEE Transactions on Dependable and Secure Computing*, vol. 5, no. 1, pp. 22-36, 2008.
- [58] R. Frampton, J. King, J. Bills and P. Wahjudi, "Marshall University cyber-security research and education system (MU CRES)," *Proceedings of the 2011 IEMS Conference*, pp. 265-269, 2011.
- [59] S. V. Radhakrishnan, A. S. Uluagac and R. Beyah, "GTID: A Technique for Physical Device and Device Type Fingerprinting," *IEEE Transactions on Dependable and Secure Computing*, pp. 1-15, 2013.
- [60] A. S. Uluagac, S. V. Radhakrishnan, C. Corbett, A. Baca and R. Beyah, "A Passive Technique for Fingerprinting Wireless Devices with Wired-side Observations," *IEEE Conference on Communications and Network Security (CNS)*, pp. 305-313, 2013.
- [61] J. Moore, D. Ali, J. Kackley and P. Wahjudi, "Analyzing the various level of security threats within the Android mobile operating system," *Proceedings of the 2011 IEMS Conference*, pp. 238-243, 2011.
- [62] Cisco Systems Inc., *Internetworking Technology Handbook*., 2009.

- [63] ESCOTAL.COM, "osi 7 layer model," 2012. [Online]. Available: <http://www.escotal.com/osilayer.html>. [Accessed 14 Aug. 2013].
- [64] L. W. Couch, Digital and Analog Communication Systems, 4 ed., New York: MacMillan, 1993.
- [65] P. J. Denning, "The Science of Computing: Computer networks," *American scientist*, vol. 73, no. 2, pp. 127-129, 1985.
- [66] B. Danev, D. Zanetti and S. Capkun, "On physical-layer identification of wireless devices," *ACM Computing Surveys*, vol. 45, no. 1, 2012.
- [67] A. King and P. Wahjudi, "Dynamic Free Text Keystroke Biometrics System for Simultaneous Authentication and Adaptation to User's Typing Pattern," *Journal of Management & Engineering Integration*, vol. 6, no. 2, pp. 86-93, 2013.
- [68] D. W. Baker, S. I. Brothers, Z. J. Geradts, D. S. Lacey, K. L. Nance, D. J. Ryan, J. E. Sammons and P. Stephenson, "Digital evolution: History, challenges and future directions for the digital and multimedia sciences section," *Forensic science: Current issues, future directions*, pp. 252-291, 2013.
- [69] A. K. Jain, A. Ross and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4-20, 2004.
- [70] D. M. Ryer, T. J. Bihl, K. W. Bauer and S. K. Rogers, "QUEST hierarchy for hyperspectral face recognition," *Advances in Artificial Intelligence*, 2012.
- [71] D. Grau, L. Zeng and Y. Xiao, "Automatically tracking engineered components through shipping and receiving processes with passive identification technologies," *Automation in Construction*, vol. 28, pp. 36-44, 2012.
- [72] M. Majzoobi, F. Koushanfar and M. Potkonjak, "Techniques for design and implementation of secure reconfigurable PUFs," *ACM Transactions on Reconfigurable Technology and Systems (TRETs)*, vol. 2, no. 1, pp. 1-33, 2009.
- [73] G. DeJean and D. Kirovski, "RF-DNA: Radio-Frequency Certificates of Authenticity," *Cryptographic Hardware and Embedded Systems (CHES)*, pp. 346-363, 2007.

- [74] W. C. Suski, M. A. Temple, M. J. Mendenhall and R. F. Mills, "Using spectral fingerprinting to improve wireless network security," *IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1-5, 2008.
- [75] K. J. Ellis and N. Serinken, "Characteristics of radio transmitter fingerprints," *Radio Science*, vol. 36, no. 4, pp. 585-597, 2001.
- [76] P. Scanlon, I. O. Kennedy and Y. Liu, "Feature extraction approaches to RF fingerprinting for device identification in femtocells," *Bell Labs Technical Journal*, vol. 15, no. 3, pp. 141-151, 2010.
- [77] T. J. Carbino, M. Temple and T. J. Bihl, "Ethernet card discrimination using unintentional cable emissions and constellation-based fingerprinting," *International Conference on Computing, Networking and Communications (ICNC)*, pp. 369-373, 2015.
- [78] B. Danev and S. Kapkun, "Transient-based identification of wireless sensor nodes," *ACM/IEEE International Conference on Information Processing in Sensor Networks*, 2009.
- [79] B. Danev, H. Luecken, S. Capkun and E. D. K., "Attacks on Physical-layer Identification," *ACM International Conference on Wireless Network Security*, 2010.
- [80] B. Danev, T. Heydt-Benjamin and S. Capkun, "Physical-layer Identification of RFID Devices," *18th Conference on USENIX Security Symposium*, pp. 199-214, 2009.
- [81] L. Desmond, C. Yuan, T. Pheng and R. Lee, "Identifying Unique Devices Through Wireless Fingerprinting," *ACM conference on Wireless Network Security*, pp. 46-55, 2008.
- [82] D. Faria and D. Cheriton, "Detecting Identity-Based Attacks In Wireless Networks Using Signalprints," *ACM Workshop on Wireless Security*, pp. 43-52, 2006.
- [83] W. Gardner, "Signal Interception: A Unifying Theoretical Framework for Feature Detection," *IEEE Transactions on Communications*, vol. 36, no. 8, pp. 897-906, 1988.
- [84] E. Hall, J. Budinger, R. Diamond and R. Apaza, "Aeronautical Mobile Communications System Development Status," *International*

Communications, Navigation and Surveillance Conference (ICNS), 2010.

- [85] J. Hall, M. Barbeau and E. Kranakis, "Detection of Transient in Radio Frequency Fingerprinting Using Signal Phase," *IASTED International Conference on Wireless and Optical Communications*, 2003.
- [86] J. Hall, M. Barbeau and E. Kranakis, "Detecting Rogue Devices in Bluetooth Networks Using Radio Frequency Fingerprinting," *Communications and Computer Networks*, pp. 108-113, 2006.
- [87] J. Hall, M. Barbeau and E. Kranakis, "Using Transceiverprints for Anomaly Based Intrusion Detection," *IASTED International Conference on Communications, Internet and Information Technology*, 2004.
- [88] P. K. Harmer and M. A. Temple, "An Improved LFS Engine for Physical Layer Security Augmentation in Cognitive Networks," *IEEE International Conference on Computing, Networking, and Communications (ICNC13)*, Jan. 2013.
- [89] C. K. Dubendorfer, B. W. Ramsey and M. A. Temple, "An RF-DNA verification process for ZigBee networks," *2012 Military Communications Conference*, pp. 1-6, 2012.
- [90] W. E. Cobb, Exploitation of Unintentional Information Leakage from Integrated Circuits, Air Force Institute of Technology: PhD Dissertation, 2011.
- [91] C. K. Dubendorfer, Using RF-DNA Fingerprints to Discriminate ZigBee Devices in an Operational Environment, Air Force Institute of Technology: MS Thesis, 2013.
- [92] P. K. Harmer, D. R. Reising and M. A. Temple, "Classifier Selection for Physical Layer Security Augmentation in Cognitive Radio Networks," *IEEE International Conference on Communications (ICC13)*, Jun. 2013.
- [93] P. K. Harmer, M. A. Temple, M. A. Buckner and E. Farquhar, "4G Security Using Physical Layer RF-DNA with DE-Optimized LFS Classification," *Journal of Communications Networks*, vol. 6, no. 9, pp. 671-681, 2011.
- [94] P. K. Harmer, M. A. Temple, M. A. Buckner and E. Farquhar, "Using Differential Evolution to Optimize 'Learning from Signals' and Enhance Network Security," *Genetic and Evolutionary Computation Conference (GECCO11)*,

July 2011.

- [95] R. Hippenstiel and Y. Payal, "Wavelet Based Transmitter Identification," *International Symposium on Signal Processing and Its Applications*, pp. 740-742, 1996.
- [96] S. Jana and S. Kasera, "Wireless Device Identification with Radiometric Signatures," *ACM International Conference on Mobile Computing and Networking (MOBICOM)*, 2008.
- [97] R. Klein, M. Temple and M. Mendenhall, "Application of wavelet-based RF fingerprinting to enhance wireless network security," *Journal of Communications and Networks*, vol. 11, no. 6, pp. 544-555, 2009.
- [98] D. R. Reising and M. A. Temple, "WiMAX Mobile Subscriber Verification Using Gabor-Based RF-DNA Fingerprints," *IEEE International Conference on Communicatiosn (ICC12)*, Jun. 2012.
- [99] D. Reising, M. Temple and J. Jackson, "Authorized and Rogue Device Discrimination Using Dimensionally Reduced RF-DNA Fingerprints," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 6, pp. 1180-1192, 2015.
- [100] D. Reising, M. Temple and J. Jackson, "Dimensionally Efficient ID Verification of OFDM-Based Devices Using GRLVQI Processing," *IEEE Journal on Selected Areas in Communications*, 2013.
- [101] M. Williams, M. Temple and D. Reising, "Augmenting big-level network security using physical layer RF-DNA fingerprinting," *IEEE Global Telecommunications Conference (GLOBECOM)*, 2010.
- [102] S. U. Rehman, K. W. Sowerby, S. Alam and I. Ardekani, "Radio frequency fingerprinting and its challenges," *IEEE Conference on Communications and Network Security (CNS)*, pp. 496-497, 2014.
- [103] S. U. Rehman, K. W. Sowerby and C. Coghill, "Analysis of impersonation attacks on systems using RF fingerprinting and low-end receivers," *Journal of Computer and System Sciences*, vol. 80, no. 3, pp. 591-601, 2014.
- [104] S. Stone and M. Temple, "Radio-frequency-based anomaly detection for programmable logic controllers in the critical infrastructure," *International Journal of Critical Infrastructure Protection*, vol. 5, no. 2, pp. 66-73, 2012.

- [105] M. Williams, S. Munns, M. Temple and M. Mendenhall, "RF-DNA fingerprinting for airport WiMax communications and security," *4th International Conference on Network and System Security (NSS)*, pp. 32-39, 2010.
- [106] J. Pang, B. Greenstein, R. Gummadi, S. Seshan and D. Wetherall, "802.11 User Fingerprinting," *International Conference on Mobile Computing and Networking (MOBICOM)*, pp. 99-110, 2007.
- [107] D. Takahashi, Y. Xiao, Y. Zhang, P. Chatzimisios and H. Chen, "IEEE 802.11 User Fingerprinting And Its Applications For Intrusion Detection," *Computers & Mathematics with Applications*, vol. 60, pp. 307-318, 2010.
- [108] O. Tekbas, O. Ureten and N. Serinken, "Improvement of Transmitter Identification System for Low SNR Transients," *IEE Electronics Letters*, vol. 40, no. 3, pp. 182-183, 2004.
- [109] D. Zanetti, B. Danev and S. Capkun, "Physical-layer Identification of UHF RFID Tags," *International Conferenece on Mobile Computing and Networking (MOBICOM)*, pp. 353-364, 2010.
- [110] J. Toonstra and W. Kinsner, "A Radio Transmitter Fingerprinting System ODO-1," *Canadian Conference on Electrical and Computer Engineering*, pp. 60-63, 1996.
- [111] J. Toonstra and W. Kinsnew, "Transient Analysis and Genetic Algorithms for Classification," *IEEE Conference on Communications, Power and Computing (WESCANEX)*, 1995.
- [112] O. Ureten and N. Serinken, "Wireless security through RF fingerprinting," *Canadian Journal of Electrical and Computer Engineering*, vol. 32, no. 1, pp. 27-33, 2007.
- [113] C. K. Dubendorfer, B. W. Ramsey and M. A. Temple, "ZigBee device verification for securing industrial control and building automation systems," *Int. Conf. on Critical Infrastructure Protection (IFIP13)*, vol. 417, pp. 47-62, 2013.
- [114] J. Hall, M. Barbeau and E. Kranakis, "Enhancing intrusion detection in wireless networks using radio frequency fingerprinting," *Communications, Internet, and Information Technology*, pp. 201-206, 2004.
- [115] Y.-J. Yuan, Z. Huang and Z.-C. Sha, "Specific Emitter Identification Based on Transient Energy Trajectory," *Progress In Electromagnetics Research C*,

vol. 44, pp. 67-82, 2013.

- [116] R. D. Deppensmith and S. J. Stone, "Integrated circuit (IC) aging effects on radio-frequency distinct native attributes (RF-DNA)," *IEEE National Aerospace and Electronics Conference (NAECON)*, pp. 331-333, 2014.
- [117] R. D. Deppensmith and S. J. Stone, "Optimized fingerprint generation using unintentional emission radio-frequency distinct native attributes (RF-DNA)," *IEEE National Aerospace and Electronics Conference (NAECON)*, pp. 327-330, 2014.
- [118] D. A. Knox and T. Kunz, "Practical RF fingerprints for Wireless Sensor Network authentication," *International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp. 531-536, 2012.
- [119] P. K. Harmer, M. D. Williams and M. A. Temple, "Using DE-Optimized LFS Processing to Enhance 4G Communication Security," *International Conference on Computer Communication Networks (ICCCN11)*, Aug. 2011.
- [120] W. Suski, M. Temple, M. Mendenhall and R. Mills, "RF Fingerprinting Commercial Communication Devices to Enhance Electronic Security," *International Journal of Electronic Security and Digital Forensics*, vol. 1, no. 3, pp. 301-322, 2008.
- [121] B. W. Ramsey, M. A. Temple and B. E. Mullins, "PHY foundation for multi-factor ZigBee node authentication," *IEEE Global Communications Conference*, pp. 795-800, 2012.
- [122] W. M. Lowder, Real-time RF-DNA fingerprinting of ZigBee devices using a software-defined radio with FPGA processing, Air Force Institute of Technology: MS Thesis, 2015.
- [123] H. J. Patel and B. W. Ramsey, "Comparison of Parametric and Non-Parametric Statistical Features for Z-Wave Fingerprinting," *IEEE Military Communications Conference (MILCOM)*, pp. 378-382, 2015.
- [124] A. B. Tao, Radio frequency distinct native attribute (RF-DNA) fingerprinting applied to SatCom short burst data modems using a software-defined radio, Air Force Institute of Technology: MS Thesis, 2015.

- [125] D. Reising, M. Temple and M. Mendenhall, "Improving Intra-Cellular Security Using Air Monitoring with RF Fingerprints," *Wireless Communications & Networking Conference (WCNC)*, 2010.
- [126] T. J. Carbino, M. A. Temple and J. Lopez, "A Comparison of PHY-Based Fingerprinting Methods Used to Enhance Network Access Control," *ICT Systems Security and Privacy Protection*, pp. 204-217, 2015.
- [127] T. Carbino, Exploitation of Unintentional Ethernet Cable Emissions Using Constellation Based-Distinct Native Attribute (CB-DNA) Fingerprints to Enhance Node Security, Air Force Institute of Technology: PhD Dissertation, 2015.
- [128] R. W. Klein, M. A. Temple, M. J. Mendenhall and D. R. Reising, "Sensitivity Analysis of Burst Detection and RF Fingerprinting Classification Performance," *IEEE International Conference on Communications (ICC09)*, pp. 1-5, Jun. 2009.
- [129] L. Wittgenstein, *Tractatus logico-philosophicus*, 1971 ed., London: Routledge & Kegan Paul, 1921.
- [130] H. Kirkham, "A Conceptual Framework for Measurement," Pacific Northwest National Laboratory, Richland, WA, 2015.
- [131] W. Heisenberg, "Physics and Philosophy: The Revolution in Modern Science," *Lectures delivered at University of St. Andrews, Scotland*, 1958.
- [132] H. J. Patel, M. Temple and R. O. Baldwin, "Improving ZigBee Device Network Authentication Using Ensemble Decision Tree Classifiers With Radio Frequency Distinct Native Attribute Fingerprinting," *IEEE Transactions on Reliability*, vol. 64, no. 1, pp. 221-233, 2015.
- [133] P. K. Harmer, Development of a Learning from Signals Classifier for Cognitive Software Defined Radio Applications, Wright Patterson AFB, OH: Ph Dissertation (DRAFT), Air Force Institute of Technology, 2013.
- [134] H. J. Patel, Advances in SCA and RF-DNA Fingerprinting Through Enhanced Linear Regression Attacks and Application of Random Forest Classifiers, Air Force Institute of Technology: PhD Dissertation, 2014.
- [135] T. J. Bihl, K. Bauer and M. A. Temple, "A Comparison of Feature Selection Methods for RF-DNA Fingerprinting Using ZigBee Device Emissions,"

IEEE Transactions on Information Forensics and Security, p. resubmitted, 2015.

- [136] W. D. Kelton, "A tutorial on design and analysis of simulation experiments," *Winter Simulation Conference*, pp. 24-31, 1995.
- [137] T. J. Bihl, M. A. Temple, K. W. Bauer and T. Paciencia, "Distance measure and algorithmic optimization approaches for learning vector quantization networks," *Neurocomputing*, p. in preparation, 2016.
- [138] L. Frenzel, "What's The Difference Between IEEE 802.15.4 And ZigBee Wireless?," 22 Mar. 2013. [Online]. Available: <http://electronicdesign.com/what-s-difference-between/what-s-difference-between-ieee-802154-and-zigbee-wireless>. [Accessed 11 Nov. 2014].
- [139] IEEE, "Overview and Guide to the IEEE 802 LMSC," Institute of Electrical and Electronics Engineers, 2004.
- [140] R. F. Heile, P. Kinney, R. Alfvén and J. Gilb, "IEEE 802.15 Wireless Personal Area Networks (WPANs) Operations Manual," Institute of Electrical and Electronics Engineers, 2013.
- [141] J.-h. Zhao, Y.-f. Li and C.-x. Xu, "Introduction of ZigBee technology," *Telecommunications for Electric Power System*, vol. 27, no. 165, pp. 54-56, 2006.
- [142] H. Kdouh, G. Zaharia, C. Brousseau, G. El Zein and G. Grunfelder, "ZigBee-based sensor network for shipboard environments," *10th International Symposium on Signals, Circuits and Systems (ISSCS)*, pp. 1-4, 2011.
- [143] D. Geer, "Users make a beeline for ZigBee sensor technology," *Computer*, vol. 38, no. 12, pp. 16-19, 2005.
- [144] T.-S. Chen, X.-G. Li, K.-H. Hsieh, Y. Xiao and P. I. Shang-Hui, "Remote doorbell control system and related smart doorbell device". US Patent 14/038,787, 27 Sept. 2013.
- [145] S. P. Kasmir and J. F. Scalisi, "Doorbell chime systems and methods". US Patent 9,160,987, 13 Oct. 2015.
- [146] W. Guo, W. M. Healy and M. Zhuo, "ZigBee-wireless mesh networks for building automation and control," *International Conference on Networking, Sensing*

and Control (ICNSC), pp. 731-736, 2010.

- [147] Z. Yiming, Y. Xianglong, G. Xishan, Z. Mingang and W. Liren, "A design of greenhouse monitoring & control system based on ZigBee wireless sensor network," *International Conference on Wireless Communications, Networking and Mobile Computing (WinCom)*, pp. 2563-2567, 2007.
- [148] P. Kadar, "ZigBee controls the household appliances," *International Conference on Intelligent Engineering Systems (INES)*, pp. 189-192, 2009.
- [149] J. S. Choi and M. Zhou, "Performance analysis of ZigBee-based body sensor networks," *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 2427-2433, 2010.
- [150] J. S. Choi and M. Zhou, "Design issues in ZigBee-based sensor network for healthcare applications," *IEEE International Conference on Networking, Sensing and Control (ICNSC)*, pp. 237-243, 2012.
- [151] S.-W. Luan, J.-H. Teng, S.-Y. Chan and L.-C. Hwang, "Development of a smart power meter for AMI based on ZigBee communication," *International Conference on Power Electronics and Drive Systems (PEDS)*, pp. 661-665, 2009.
- [152] C. Gezer and C. Buratti, "A ZigBee smart energy implementation for energy efficient buildings," *IEEE 73rd Vehicular Technology Conference (VTC Spring)*, pp. 1-5, 2011.
- [153] N. C. Batista, R. Melicio, J. C. Matias and J. P. Catalao, "ZigBee standard in the creation of wireless networks for advanced metering infrastructures," *16th IEEE Mediterranean Electrotechnical Conference (MELECON)*, pp. 220-223, 2012.
- [154] C. Bennett and D. Highfill, "Networking AMI smart meters," *IEEE Energy 2030 Conference*, pp. 1-8, 2008.
- [155] M. Anas, N. Javaid, A. Mahmood, S. M. Raza, U. Qasim and Z. A. Khan, "Minimizing electricity theft using smart meters in AMI," *Seventh International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, pp. 176-182, 2012.
- [156] S. S. Depuru, L. Wang and V. Devabhaktuni, "Smart meters for power grid: Challenges, issues, advantages and status," *Renewable and Sustainable*

Energy Reviews, vol. 15, no. 6, pp. 2736-2742, 2011.

- [157] G. Zheng, H. Zong and Z. Zhang, "A scheme of telemetering and preventing electric larceny system based on GPRS communication system," *7th World Congress on Intelligent Control and Automation*, pp. 5148-5151, 2008.
- [158] D.-M. Han and J.-H. Lim, "Smart home energy management system using IEEE 802.15. 4 and zigbee," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 3, pp. 1403-1410, 2010.
- [159] M. Erol-Kantarci and H. T. Mouftah, "Wireless sensor networks for cost-efficient residential energy management in the smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 314-325, 2011.
- [160] J. Sun, N. Wang and L. Liu, "A novel water-sludge interface data transfer system applying ZigBee wireless network," *International Journal of Information Acquisition*, vol. 3, no. 4, pp. 301-309, 2006.
- [161] K. N. Kumar, R. Prabakaran, V. Sarma Dhulipala and P. Ranjith, "Future sensors and utilization of sensors in chemical industries with control of environmental hazards," *International Conference on Environmental Science and Development*, pp. 224-228, 2011.
- [162] D. Nowak, L. Krzak and C. Worek, "Integration of ZigBee and IEC 61850 networks for a substation automation system," *IEEE/PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, pp. 1-5, 2013.
- [163] B. Chen, M. Wu, S. Yao and N. Binbin, "ZigBee Technology and Its Application on Wireless Meter-reading System," *IEEE International Conference on Industrial Informatics*, pp. 1257-1260, 2006.
- [164] A. Biswas, A. Alkhalid, T. Kunz and C.-H. Lung, "A lightweight defence against the packet in packet attack in ZigBee networks," *IFIP Wireless Days*, pp. 1-3, 2012.
- [165] B. Danev, H. Luecken, S. Capkun and K. El Defrawy, "Attacks on physical-layer identification," *Proceedings of the 3rd ACM conference on wireless network security*, pp. 89-98, 2010.
- [166] L. Xiao, L. Greenstein, N. Mandayam and W. Trappe, "Fingerprints in the ether: Using the physical layer for wireless authentication," *IEEE International*

Conference on Communications, pp. 4646-4651, 2007.

- [167] Texas Instruments, "2.4 GHz IEEE 802.15.4 / ZigBee-ready RF Transceiver," Texas Instruments, 2014.
- [168] S. Farahani, *ZigBee Wireless Networks and Transceivers*, Newnes Press, 2011.
- [169] M. Hazen, "Z-Wave vs. ZigBee—What's the difference," *Mobile Dev & Design*, 25 July 2006.
- [170] M. Knight, "How safe is Z-Wave?[Wireless standards]," *Computing and Control Engineering*, vol. 17, no. 6, pp. 18-23, 2006.
- [171] S. Omatseye, "Z battle for Z wirelessly controlled home," *RCR News*, vol. 24, no. 5, 2005.
- [172] C. Gomez and J. Paradells, "Wireless home automation networks: A survey of architectures and technologies," *IEEE Communications Magazine*, pp. 92-101, June 2010.
- [173] M. Galeev, "Catching the Z-Wave," *Electronic Engineering Times India*, pp. 1-5, Oct. 2006.
- [174] Aeon Labs, "Aeon Labs Z-Stick Series 2," Aeon Labs.
- [175] A. A. Tomko, C. J. Rieser and L. H. Buell, "Physical-layer intrusion detection in wireless networks," *IEEE Military Communications Conference (MILCOM)*, pp. 1-7, 2006.
- [176] F. Holker, T. Moss, B. Griefahn, W. Kloas, C. C. Voight, D. Henckel, A. Hanel, P. M. Kappeler, S. Volker, A. Schwöpe, S. Franke, D. Uhrlandt, J. Fischer, R. Klenke, C. Wolter and K. Tockner, "The dark side of light: A transdisciplinary research agenda for light pollution policy," *Ecology and Society*, vol. 15, no. 4, 2010.
- [177] F. Leferink, F. Silva, J. Catrysse, S. Batterman, V. Beauvois and A. Roc'h, "Man-made noise in our living environments," *Radio Science Bulletin*, no. 334, pp. 49-57, 2010.
- [178] S. Pennesi and S. Sebastiani, "Information security and emissions control," *International Symposium on Electromagnetic Compatibility*, vol. 3, pp. 777-

781, 2005.

- [179] T. S. Perry and L. Geppert, "Do portable electronics endanger flight? The evidence mounts," *IEEE Spectrum*, vol. 33, no. 9, pp. 26-33, 1996.
- [180] M. Mardiguian, *Controlling Radiated Emissions by Design*, Springer, 2001.
- [181] B. Archambeault, "Proper design of intentional splits in the ground reference plane of PC boards to minimize emissions on I/O wires and cables," *IEEE International Symposium on Electromagnetic Compatibility*, vol. 2, pp. 768-773, 1998.
- [182] M. I. Montrose, *EMC and the Printed Circuit Board: Design, Theory, and Layout Made Simple*, John Wiley & Sons, 2004.
- [183] D. Agrawal, B. Archambeault, J. R. Rao and P. Rohatgi, "The EM Side—Channel(s)," *Cryptographic Hardware and Embedded Systems - CHES 2002*, vol. 2523, pp. 29-45, 2003.
- [184] H. Weng, X. Dong, X. Hu, D. G. Beetner, T. Hubing and D. Wunsch, "Neural network detection and identification of electronic devices based on their unintended emissions," *International Symposium on Electromagnetic Compatibility*, vol. 1, pp. 245-249, 2005.
- [185] C. M. Roberts, "Radio frequency identification (RFID)," *Computers & Security*, vol. 25, no. 1, pp. 18-26, 2006.
- [186] J. Landt, "The history of RFID," *IEEE Potentials*, vol. 24, no. 4, pp. 8-11, 2005.
- [187] W. S. Holland, W. A. Young and G. R. Weckman, "Facility RFID localization system based on artificial neural networks," *International Journal of Industrial Engineering: Theory, Applications and Practice*, vol. 18, no. 1, pp. 16-24, 2011.
- [188] M. Weber, U. Birkel, R. Collmann and J. Engelbrecht, "Comparison of various methods for indoor RF fingerprinting using leaky feeder cable," *Workshop on Positioning Navigation and Communication (WPNC)*, pp. 291-298, 2010.
- [189] A. Candore, O. Kocabas and F. Koushanfar, "Robust stable radiometric fingerprinting for wireless devices," *IEEE International Workshop on*

Hardware-Oriented Security and Trust (HOST), pp. 43-49, 2009.

- [190] R. W. Klein, Application of dual-tree complex wavelet transforms to burst detection and RF fingerprint classification, Air Force Institute of Technology: PhD Dissertation, 2009.
- [191] A. L. Lance, W. D. Seal and F. Labaar, "Phase noise and AM noise measurement in the frequency domain," *Infrared and Millimeter Waves*, vol. 11, pp. 239-289, 1984.
- [192] D. E. Vakman and L. A. Vainshtein, "Amplitude, phase, frequency - fundamental concepts of oscillation theory," *Soviet Physics Uspekhi*, vol. 20, no. 12, pp. 1002-1016, 1977.
- [193] P. Diaconis, "Application of the method of moments in probability and statistics," in *Moments in Mathematics*, SIAM, 1987.
- [194] A. M. Fiori and M. Zenga, "Karl Pearson and the Origin of Kurtosis," *International Statistical Review*, vol. 77, no. 1, pp. 40-50, 2009.
- [195] D. E. Scates, "Characteristics of Kurtosis," *The Journal of Experimental Education*, vol. 11, no. 3, pp. 226-237, 1943.
- [196] W. Kim, F. Fabozzi, P. Cheridito and C. Fox, "Controlling portfolio skewness and kurtosis without directly optimizing third and fourth moments," *Economics Letters*, vol. 122, no. 2, pp. 154-158, 2014.
- [197] B. Aracıoğlu, F. Demircan and H. Soyuer, "Mean-Variance-Skewness-Kurtosis Approach to Portfolio Optimization: An Application in İstanbul Stock Exchange.," *Ege Academic Review*, vol. 11, pp. 9-17, 2011.
- [198] L. You, "Higher order moment risk in efficient futures portfolios," *Journal Of Economics And Business*, vol. 65, no. 2, pp. 33-54, 2013.
- [199] S. Thirukkonda, Correlation in Firm Default Behavior, M.S. Thesis: Massachusetts Institute of Technology, 2009.
- [200] M. Laurin, "Assessment of the Relative Merits of a Few Methods to Detect Evolutionary Trends," *Systematic Biology*, vol. 59, no. 6, pp. 689-704, 2010.

- [201] V. Lohweg, J. L. Hoffmann, H. Dörksen, R. Hildebrand, E. Gillich, J. Hofmann and J. Schaed, "Banknote authentication with mobile devices," *IS&T/SPIE Electronic Imaging*, pp. 1-14, 2013.
- [202] "Descriptive Statistics," in *Biostatistics*, Wiley, 2013, pp. 19-64.
- [203] J. Barnes, *The Presocratic Philosophers*, Routledge, 1983.
- [204] K. Freeman, *Ancilla to the Pre-Socratic Philosophers*, Harvard University Press , 1948.
- [205] J. Burnet, *Early Greek Philosophy*, A. and C. Black, 1892.
- [206] Aristotle, "Categories," in *The Complete Works of Aristotle*, Princeton University Press, 1995, pp. 3-24.
- [207] H.-C. Ni, *The Complete Works of Lao Tzu: Tao Teh Ching & Hau Hu Ching*, Sevenstar Communications, 1995.
- [208] R. L. Fry, *Personal Communication*, AFIT Colloquium, 2012.
- [209] D. Odegard, "Locke and mind-body dualism," *Philosophy*, vol. 45, no. 172, pp. 87-105, 1970.
- [210] S. Tomaselli, "The First Person: Descartes, Locke and Mind-Body Dualism," *History of Science*, vol. 22, pp. 185-205, 1984.
- [211] D. Hume, *A Treatise of Human Nature*, London, 1738.
- [212] P. E. Ross, "As I See It: Flash of genius," *Forbes*, pp. 98-104, Nov. 1998.
- [213] A. K. Jain, R. P. Duin and J. Mao, "Statistical Pattern Recognition: a Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. vol. 22, no. no. 1, pp. 4-37, Jan. 2000.
- [214] J. F. F. von Bielfeld, "Chap. XIII: Statistics," in *The Elements of Universal Erudition: Containing an Analytical Abridgment of the Sciences, Polite Arts, and Belles Lettres, Volume 3*, G. Scott, 1770, pp. 268-279.
- [215] K. S. Sidhu, "Introduction," in *Statistics In Education And Psychology*, Sterling Publishers Pvt. Ltd, 2006, pp. 1-9.

- [216] R. O. Duda, P. E. Hart and D. G. Stork, Pattern Classification, 2nd ed., John Wiley & Sons, 2001.
- [217] B. K. Williams, "Some observations of the use of discriminant analysis in ecology," *Ecology*, vol. 64, no. 5, pp. 1283-1291, 1983.
- [218] S. Manel, J.-M. Dias and S. J. Ormerod, "Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird," *Ecological Modelling*, vol. 120, pp. 337-347, 1999.
- [219] U. Jumhawan, S. P. Putri, Yusianto, E. Marwani, T. Bamba and E. Fukusaki, "Selection of discriminant markers for authentication of Asian Palm Civet Coffee (Kopi Luwak): A metabolomics approach," *Journal of Agricultural and Food Chemistry*, vol. 61, no. 33, pp. 7994-8001, 2013.
- [220] C. J. Huberty and M. H. Hussein, "Some problems in reporting use of discriminant analysis," *The Journal of Experimental Education*, vol. 71, no. 2, pp. 177-191, 2003.
- [221] P. Shin and K. Fong, "Multiple discriminant analysis of marine sediment data," *Marine Pollution Bulletin*, vol. 39, no. 1-12, pp. 285-294, 1999.
- [222] L. A. Le Blanc and C. T. Rucks, "A multiple discriminant analysis of vessel accidents," *Accident Analysis & Prediction*, vol. 28, no. 4, pp. 501-510, 1996.
- [223] B. J. Steward, K. W. Bauer and G. P. Perram, "Remote discrimination of large-caliber gun firing signatures," *Journal of Applied Remote Sensing*, vol. 6, p. 16, 2012.
- [224] W. E. Cobb, E. D. Laspe, R. O. Baldwin, M. A. Temple and Y. Kim, "Intrinsic Physical-Layer Authentication of Integrated Circuits," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 1, pp. 14-24, Feb. 2012.
- [225] S. J. Dixon and R. G. Brereton, "omparison of performance of five common classifiers represented as boundary methods: Euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on," *Chemometrics and Intelligent Laboratory Systems*, vol. 95, no. 1, pp. 1-17, 2009.

- [226] P.-W. Lei and L. M. Koehly, "Linear discriminant analysis versus logistic regression: A comparison of classification errors in the two-group case," *The Journal of Experimental Education*, vol. 72, no. 1, pp. 25-49, 2003.
- [227] S. MacDonell and A. Gray, "Software forensics applied to the task of discriminating between program authors," *Journal of Systems Research and Information Systems*, vol. 10, pp. 113-127, 2001.
- [228] V. S. Desai, J. N. Crook and G. A. Overstreet, "A comparison of neural networks and linear scoring models in the credit union environment," *European Journal of Operations Research*, vol. 95, pp. 24-37, 1996.
- [229] T. Li, S. Zhu and M. Ogihara, "Using discriminant analysis for multi-class classification: an experimental investigation," *Knowledge and Information Systems*, vol. 10, no. 4, pp. 453-472, 2006.
- [230] M. Bressan and J. Vitria, "Nonparametric discriminant analysis and nearest neighbor classification," *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2743-2749, 2003.
- [231] X. Zhan and B. Ma, "Kernel Nonparametric Discriminant Analysis," *International Conference on Electrical and Control Engineering (ICECE)*, pp. 4544-4547, 2011.
- [232] Z. Gu and J. Yang, "A new nonparametric linear discriminant analysis method based on marginal information," *Chinese Conference on Pattern Recognition (CCPR)*, pp. 1-5, 2009.
- [233] Q. Zeng and C. Wang, "NPDA/CS: Improved non-parametric discriminant analysis with CS decomposition and its application to face recognition," *IEEE International Conference on Image Processing (ICIP)*, pp. 4537-4540, 2010.
- [234] T. Hastie, R. Tibshirani and A. Buja, "Flexible discriminant analysis by optimal sorting," *Journal of the American Statistical Society*, vol. 89, no. 428, pp. 1255-1270, 1994.
- [235] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annual Eugenics*, vol. 7, pp. 179-188, 1936.
- [236] C. R. Rao, "The utilization of multiple measurements in problems of biological classification," *Journal of the Royal Statistical Society, Series B*

(*Methodological*), vol. 10, no. 2, pp. 159-203, 1948.

- [237] W. R. Dillon and M. Goldstein, *Multivariate Analysis Methods and Applications*, New York: John Wiley & Sons, 1984.
- [238] J. W. Sammon, "An optimal discriminant plane," *IEEE Transactions on Computers*, Vols. C-19, no. 9, pp. 826-829, 1970.
- [239] E. R. Pinch, *Optimal Controls and the Calculus of Variations*, Oxford University Press, 1995.
- [240] M. Welling, *Fisher linear discriminant analysis*, Department of Computer Science: University of Toronto, 2005.
- [241] B. W. Ramsey, *Improved Wireless Security Through Physical Layer Protocol Manipulation and Radio Frequency Fingerprinting*, PhD Dissertation: Air Force Institute of Technology, 2014.
- [242] A. Ferreira, S. Chen, M. A. Simaan, J. R. Boston and J. F. Antaki, "A discriminant-analysis-based suction detection system for rotary blood pumps," *IEEE EMBS Annual International Conference*, pp. 5382-5385, 2006.
- [243] K. W. Bauer, "OPER 661 and OPER 685 Lecture Notes," Air Force Institute of Technology, Wright Patterson AFB, OH, 2014.
- [244] M. J. Mendenhall, *A Neural Relevance Model for Feature Extraction from Hyperspectral Images, and its Application in the Wavelet Domain*, PhD Dissertation: Rice University, 2006.
- [245] M. J. Mendenhall and E. Merenyi, "Relevance-Based Feature Extraction for Hyperspectral Imagery," *IEEE Transactions on Neural Networks*, vol. 19, no. 4, pp. 658-672, 2008.
- [246] M. Mendenhall and E. Merenyi, "Relevance-based feature extraction from hyperspectral images in the complex wavelet domain," *IEEE Mountain Workshop on Adaptive and Learning System*, pp. 24-29, 2006.
- [247] J. A. Gonzalez, *Numerical Analysis for Relevant Features in Intrusion Detection (NARFid)*, Air Force Institute of Technology: MS Thesis, 2009.
- [248] Y.-C. Hu, Wen, Chia-Hsien, C.-C. Lo and W.-L. Chen, "Image vector quantization using geometric transform and lossless index coding," *Optical Engineering*,

vol. 52, no. 3, 2013.

- [249] B. Purnamasari, F. Chandra and S. Dyah, "Application of neural networks on blood serum image for early detection of typhus," *Indonesian Journal of Tropical and Infectious Disease*, vol. 4, no. 4, pp. 53-58, 2013.
- [250] C. G. Looney, Pattern Recognition Using Neural Networks, Oxford University Press, 1997.
- [251] R. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, vol. 3, no. 4, pp. 4-22, 1987.
- [252] D. Nova and P. A. Estevez, "A review of learning vector quantization classifiers," in *Neural Computing and Applications* , 2013, pp. 1-14.
- [253] C. Lv, X. An, Z. Liu and Q. Zhao, "Dual weight learning vector quantization," *ICSP Proceedings*, pp. 1722-1725, 2008.
- [254] A. K. Jain, J. Mao and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, pp. 31-44, 1996.
- [255] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464-1480, 1990.
- [256] T. Kohonen, Self-Organizing Maps, Berlin: Springer, 1997.
- [257] T. Kohonen, "Learning vector quantization," in *The handbook of brain theory and neural networks*, Cambridge, MA, MIT Press, 1995, pp. 537-540.
- [258] M. Kaden, M. Lange, D. Nebel, M. Riedel, T. Geweniger and T. Villmann, "Aspects in classification learning-Review of recent developments in Learning Vector Quantization," *Foundations of Computing and Decision Sciences*, vol. 39, no. 2, pp. 79-105, 2014.
- [259] P. D. Wasserman, "Counterpropagation Networks," in *Neural Computing: Theory and Practice*, van Nostrand Reinhold, 1989, pp. 61-75.
- [260] P. D. Wasserman, Neural Computing, Theory and Practics, Van Nostrand Reinhold, 1989.
- [261] P. D. Wasserman, "Appendix C: Training Algorithms," in *Neural Computing:*

Theory and Practice, van Nostrand Reinhold, 1989, pp. 211-222.

- [262] G. M. Georgiou, "Single-layer networks," in *Handbook of Neural Computation*, IOP Publishing Ltd and Oxford University Press, 1997, pp. C1.1:1-24.
- [263] R. Hecht-Nielsen, "Theory of the backpropagation neural network," *International Joint Conference on Neural Networks (IJCNN)*, pp. 593-605, 1989.
- [264] T. Bojer, B. Hammer, D. Schunk and K. Tluk von Toschanowitz, "Relevance determination in learning vector quantization," *Proceedings of European Symposium on Artificial Neural Networks (ESANN)*, pp. 271-276, 2001.
- [265] A. S. Sato and K. Yamada, "Generalized learning vector quantization," in *Advances in neural information processing systems*, Cambridge, MA, MIT Press, 1995, pp. 423-429.
- [266] B. Hammer and T. Villmann, "Generalized relevance learning vector quantization," *Neural Networks*, vol. 15, no. 8-9, pp. 1059-1068, 2002.
- [267] M.-T. Vakil-Baghmisheh and N. Pavesic, "Premature clustering phenomenon and new training algorithms for LVQ," *Pattern Recognition*, vol. 36, pp. 1901-1912, 2003.
- [268] D. Hofmann, Gisbrecht, Andrej and B. Hammer, "Efficient approximations of kernel robust soft lvq," *Advances in Self-Organizing Maps*, pp. 183-192, 2013.
- [269] A. K. Qin and P. N. Suganthan, "A Novel Kernel Prototype-Based Learning Algorithm," *International Conference on Pattern Recognition*, pp. 621-624, 2004.
- [270] K. Torkkola and W. M. Campbell, "Mutual information in learning feature transformations," *Proceedings of the 17th International Conference on Machine Learning*, pp. 1015-1022, 2000.
- [271] S. Kaski, J. Kangas and T. Kohonen, "Bibliography of self-organizing map (SOM) papers: 1981–1997," *Neural computing surveys*, vol. 1, no. 3&4, pp. 1-176, 1998.
- [272] R. M. Gray, "Vector quantization," *IEEE ASSP Magazine*, vol. 1, no. 2, pp. 4-29, 1984.

- [273] A. I. Gonzalez, M. Grana and A. D'Anjou, "An analysis of the GLVQ algorithm," *IEEE Transactions on Neural Networks*, vol. 6, no. 4, pp. 1012-1016, 1995.
- [274] P. V. (. Balakrishnan, M. C. Cooper, V. S. Jacob and P. A. Lewis, "A study of the classification capabilities of neural networks using unsupervised learning: A comparison with K-means clustering," *Psychometrika*, vol. 59, no. 4, pp. 509-525, 1994.
- [275] A. N. Albatineh and M. Niewiadomska-Bugaj, "MCS: A Method for Finding the Number of Clusters," *Journal of Classification*, vol. 28, pp. 184-209, 2011.
- [276] O. Granichin, D. Shalymov, R. Avros and Z. Volkovich, "A Randomized Algorithm for Estimating the Number of Clusters," *Automation and Remote Control*, vol. 72, no. 4, pp. 754-765, 2011.
- [277] J. Al Shaqsi and W. Wang, "Estimating the predominant number of clusters in a dataset," *Intelligent Data Analysis*, vol. 17, pp. 603-626, 2013.
- [278] A. K. Jain, M. N. Murty and P. J. Flynn, "Data clustering: a review," *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264-323, 1999.
- [279] "4.2 Increments, Differentials, and Linear Approximation," in *Calculus with Analytic Geometry*, Prentice Hall, 1998, pp. 200-206.
- [280] T. Kohonen, J. Kangas, J. Laaksonen and K. Torkkola, "LVQ_PAK: A program package for the correct application of Learning Vector Quantization algorithms," *Proceedings of the International Joint Conference on Neural Networks*, pp. 725-730, 1992.
- [281] B. Hammer, M. Strickert and T. Villmann, "On the generalization ability of GRLVQ networks".
- [282] T. Kohonen, "Improved versions of learning vector quantization," *IEEE International Joint Conference on Neural Networks (IJCNN)*, pp. 545-550, 1990.
- [283] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *International Journal of Mathematical Models and Methods in Applied Science*, vol. 4, no. 1, pp. 300-307, 2007.
- [284] T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of

the vegetation on Danish commons," *Kongelige Danske Videnskabernes Selskab*, vol. 5, no. 4, p. 1–34, 1948.

- [285] Z. Zhang, R. Deriche, O. Faugeras and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78, no. 1-2, pp. 87-119, 1995.
- [286] H. Sun, G. Skogerbo, X. Zheng, W. Liu and Y. Li, "Genomic regions with distinct genomic distance conservation in vertebrate genomes," *BMC Genomics*, vol. 10, pp. 1-12, 2009.
- [287] J. A. Paradiso and N. Gershenfeld, "Musical applications of electric field sensing," *Computer music journal*, pp. 69-89, 1997.
- [288] N. Matsuura, M. L. Hill, I. Gorelikov, S. Zhu, K. Wan, J. G. Mainprize, M. J. Yaffe and J. A. Rowlands, "Development of a targeted CT contrast agent: assessment of cellular interactions using novel integrated optical labels," *SPIE Medical Imaging*, 2009.
- [289] R. E. Crippen, "Calculating the vegetation index faster," *Remote Sensing of Environment*, vol. 34, no. 1, pp. 71-73, 1990.
- [290] M. Biehl, B. Hammer and T. Villmann, "Distance measures for prototype based classification," *Proc. International Workshop on Brain-Inspired Computing*, 2013.
- [291] M. Strickert, U. Seiffert, N. Sreenivasulu, W. Weschke, T. Villmann and B. Hammer, "Generalized relevance LVQ (GRLVQ) with correlation measures for gene expression analysis," *Neurocomputing*, vol. 69, no. 7-9, pp. 651-659, 2006.
- [292] A. A. Minai and R. D. Williams, "On the derivatives of the sigmoid," *Neural Networks*, vol. 6, no. 6, pp. 846-853, 1993.
- [293] D. DeSieno, "Adding a conscience to competitive learning," *Proceedings of the IEEE International Conference on Neural Networks*, pp. 117-124, 1988.
- [294] T.-S. Lim, W.-Y. Loh and Y.-S. Shih, "A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms," *Machine Learning*, vol. 40, pp. 203-229, 2000.

- [295] S. Bischoff, M. J. Mendenhall, A. C. Rice and J. R. Vasquez, "Adapting learning parameter transition in the generalized learning vector quantization family of classifiers," *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1-4, 2010.
- [296] J. A. Kangas, T. K. Kohonen and J. T. Laaksonen, "Variants of self-organizing maps," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 93-99, 1990.
- [297] P. L. Bartlett and S. Mendelson, "Rademacher and Gaussian complexities: Risk bounds and structural results," *The Journal of Machine Learning Research*, vol. 3, pp. 463-482, 2003.
- [298] P. Schneider, M. Biehl and B. Hammer, "Distance learning in discriminative vector quantization," *Neural Computation*, vol. 21, no. 10, pp. 2942-2969, 2009.
- [299] M. D. Mackenzie, "Enhancements to the dynamic self-organizing neural network," *Neural Computing and Applications*, vol. 4, pp. 44-57, 1996.
- [300] M.-C. Su, T.-K. Liu and H.-T. Chang, "An efficient initialization scheme for the self-organizing feature map algorithm," *Proceedings of the International Joint Conference on Neural Networks*, pp. 1906-1910, 1999.
- [301] A. Sato, "An analysis of initial state dependence in generalized LVQ," *International Conference on Artificial Neural Networks*, pp. 928-933, 1999.
- [302] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, John Wiley & Sons, 2011.
- [303] P. Thulasiraman, *Semantic Classification of Rural and Urban Images Using Learning Vector Quantization*, Louisiana State University: MS Thesis, 2005.
- [304] H. J. Gage, *Using upper layer weights to efficiently construct and train feedforward neural networks executing backpropagation*, Air Force Institute of Technology: MS Thesis, 2011.
- [305] P. Welch, *On the problem of the initial transient in steady-state simulation*, IBM Watson Research Center, 1981.
- [306] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol.

27, no. 8, pp. 861-874, 2006.

- [307] M. R. Gupta, S. Bengio and J. Weston, "Training highly multiclass classifiers," *The Journal of Machine Learning Research*, vol. 1, no. 1461-1492, p. 15, 2014.
- [308] J. C. Platt, N. Cristianini and J. Shawe-Taylor, "Large Margin DAGs for Multiclass Classification," *Neural Information Processing Systems*, vol. 12, pp. 547-553, 2000.
- [309] W. Bi and J. Kwok, "Efficient multi-label classification with many labels," *International Conference on Machine Learning*, pp. 405-413, 2013.
- [310] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao and T. Huang, "Large-scale image classification: fast feature extraction and svm training," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1689-1696, 2011.
- [311] T.-N. Doan, T.-N. Do and F. Poulet, "Large Scale Image Classification with Many Classes, Multi-features and Very High-Dimensional Signatures," *Advanced Computational Methods for Knowledge Engineering*, pp. 105-116, 2013.
- [312] O. Madani and M. Connor, "Large-Scale Many-Class Learning," *SDM*, pp. 846-857, 2008.
- [313] T.-N. Do, P. Lenca and S. Lallich, "Classifying many-class high-dimensional fingerprint datasets using random forest of oblique decision trees," *Vietnam Journal of Computer Science*, vol. 2, no. 1, pp. 3-12, 2014.
- [314] R. Saidi, W. Dhifli, S. Aridhi, M. Agier, G. Bronner, D. Debroas, L. D'Orazio, F. Enault, S. Guillaume and E. Mephu, "Protein classification in the case of large and many-class datasets," *International Conference on Biology, Informatics, and Mathematics (JOBIM)*, 2010.
- [315] Z. Akata, F. Perronnin, Z. Harchaoui and C. Schmid, "Good practice in large-scale learning for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 507-520, 2014.
- [316] J. Sánchez and F. Perronnin, "High-dimensional signature compression for large-scale image classification," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1665-1672, 2011.

- [317] M. Zhu, "The Limitations of LDA," in *Feature Extraction and Dimension Reduction with Applications to Classification and the Analysis of Co-occurrence Data*, Stanford University, PhD Dissertation, 2001, pp. 28-29.
- [318] T. Martin, "Fuzzy sets in the fight against digital obesity," *Fuzzy Sets and Systems*, vol. 156, no. 3, pp. 411-417, 2005.
- [319] T. Martin, Y. Shen and A. Majidian, "Discovery of time-varying relations using fuzzy formal concept analysis and associations," *International Journal of Intelligent Systems*, vol. 25, no. 12, pp. 1217-1248, 2010.
- [320] N. Holmes, "The rise and rise of digital gluttony," *Computer*, pp. 96, 94-95, May 2010.
- [321] V. M. Megler and D. Maier, "When big data leads to lost data," *Proceedings of the 5th Ph.D. Workshop on Information and Knowledge*, pp. 1-8, 2012.
- [322] T. Hey and A. E. Trefethen, "The data deluge: An e-science perspective," in *Grid Computing: Making the Global Infrastructure a Reality*, Chichester, UK, John Wiley & Sons, 2003, pp. 809-824.
- [323] N. Holmes, "Aspects of data obesity," *Computer*, vol. 44, no. 5, pp. 104, 102-103, 2011.
- [324] B. Ratner, *Statistical Modeling and Analysis for Database Marketing: Effective Techniques for Mining Big Data*, CRC Press, 2003.
- [325] B. M. Gross, *The Managing of Organizations: The Administrative Struggle*, 1964.
- [326] D. Rosenberg, "Early modern information overload," *Journal of the History of Ideas*, vol. 64, no. 1, pp. 1-9, 2003.
- [327] T. J. Bihl, W. A. Young and G. R. Weckman, "Defining, Understanding, and Addressing Big Data," *International Journal of Business Analytics (IJBAN)*, to appear.
- [328] J. Ruskin, *Sesame and Lilies*, New York: John Wiley & Son, 1865.
- [329] F. A. Hayek, "The Use of Knowledge in Society," *The American Economic Review*, vol. vol. 35, no. no. 4, pp. 519-530, 1945.

- [330] K. W. Bauer, S. G. Alsing and K. A. Greene, "Feature screening using signal-to-noise ratios," *Neurocomputing*, vol. 31, pp. 29-44, 2000.
- [331] L. Song, A. Smola, A. Gretton, J. Bedo and K. Borgwardt, "Feature selection via dependence maximization," *Journal of Machine Learning Research*, vol. 13, pp. 1393-1434, 2012.
- [332] S. C. H. Hoi, J. Wang, P. Zhao and R. Jin, "Online feature selection for mining big data," *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, pp. 93-100, 2012.
- [333] Y. Saeys, I. Inza and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507-2517, 2007.
- [334] A. G. K. Janecek, W. N. Gansterer, M. A. Demel and G. F. Ecker, "On the Relationship Between Feature Selection and Classification Accuracy," *JMLR: Workshop and Conference Proceedings*, vol. 4, pp. 90-105, 2008.
- [335] P. Schrater, *Feature Selection/Extraction*, University of Minnesota : Lecture Notes for CSCI 5521, 2009.
- [336] Y.-M. Chen and J.-J. Liu, "Cost-effective design for injection molding," *Robotics and Computer-Integrated Manufacturing*, vol. 15, pp. 1-21, 1999.
- [337] F. J. Massey, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68-78, 1951.
- [338] A. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *Giornale dell' Istituto Italiano degli Attuari*, vol. 4, pp. 1-11, 1933.
- [339] H. Smirnov, "Sur les Ecarts de la Courbe de Distribution Empirique," *Recueil Mathematique*, vol. 6, pp. 3-26, 1939.
- [340] W. J. Conover, *Practical Nonparametric Statistics*, 2nd ed., New York: John Wiley & Sons, 1980.
- [341] "13.7 The Kolmogorov-Smirnov Goodness-of-Fit," in *Biostatistics*, Wiley, 2013, pp. 698-703.
- [342] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical*

Recipes in C, Cambridge Univ. Press, 1992.

- [343] W. J. Conover, "A Kolmogorov Goodness-of-Fit Test for Discontinuous Distributions," *Journal of the American Statistical Association*, vol. 67, no. 339, pp. 591-596, 1972.
- [344] L. H. Miller, "Table of percentage points of Komogorov statistics," *Journal of the American Statistical Association*, vol. 51, no. 273, pp. 111-121, 1956.
- [345] M. Stephens, "Use of the Kolmogorov-Smirnov, Cramer-von Mises and related statistics without extensive tables," *Journal of the Royal Statistical Society Series B*, vol. 32, no. 1, pp. 115-122, 1970.
- [346] N. A. Nechval, "Goodness-of-fit testing in the presence of nuisance parameters with applications to feature selection and pattern recognition in digital image processing," *Visual Communications and Image Processing*, pp. 694-701, 1988.
- [347] N. A. Nechval, E. K. Vasermanis, U. Rozevskis, K. Rozite and K. N. Nechval, "Characterizations theorems and goodness-of-fit testing on small data samples," *Transport and Telecommunication*, vol. 4, no. 2, pp. 92-100, 2003.
- [348] B. W. Ramsey, M. A. Temple and B. E. Mullins, "PHY foundation for multi-factor ZigBee node authentication," *IEEE Global Communications Conference (GLOBECOM)*, pp. 795-800, 2012.
- [349] J. D. Habbema and J. Hermans, "Selection of variables in discriminant analysis by F-statistic and error rate," *Technometrics*, vol. 19, no. 4, pp. 487-493, 1977.
- [350] D. V. Raje, H. J. Purohit and R. N. Singh, "Distinguishing features of 16S rDNA gene for five dominating bacterial genus observed in bioremediation," *Journal of Computational Biology*, vol. 9, no. 6, pp. 819-829, 2002.
- [351] C. J. Huberty and J. M. Wisenbaker, "Variable importance in multivariate group comparisons," *Journal of Education Statistics*, vol. 17, no. 1, pp. 75-91, 1992.
- [352] P. Jafari and F. Azuaje, "An assessment of recently published gene expression data analyses: reporting experimental design and statistical factors," *BMC Medical Informatics and Decision Making*, vol. 6, no. 1, 2006.

- [353] M. A. Hall and L. A. Smith, "Practical feature subset selection for machine learning," *Proceedings of the 21st Australasian Computer Science Conference*, pp. 181-191, 1998.
- [354] M. H. Kutner, C. J. Nachtsheim and J. Neter, *Applied Linear Regression Models*, 4th ed., McGraw-Hill/Irwin, 2004.
- [355] H. B. Mann, *Analysis and Design of Experiments*, New York: Dover Publications, INC., 1949.
- [356] J. D. Gibbons and J. W. Pratt, "P-values: Interpretation and Methodology," *The American Statistician*, vol. 29, no. 1, pp. 20-25, 1975.
- [357] R. Nuzzo, "Statistical Errors," *Nature*, vol. 506, pp. 150-152, Feb. 2014.
- [358] T. Wu, J. Duchateau, J.-P. Martens and D. van Compernelle, "Feature subset selection for improved native accent identification," *Speech Communication*, vol. 52, no. 2, pp. 83-98, 2010.
- [359] K. Schmidt, T. Behrens and T. Scholten, "Instance selection and classification tree analysis for large spatial datasets in digital soil mapping," *Geoderma*, vol. 146, no. 1-2, pp. 138-146, 2008.
- [360] R. C. Craddock, P. E. Holtzheimer, X. P. Hu and H. S. Mayberg, "Disease state prediction from resting state functional connectivity," *Magnetic resonance in Medicine*, vol. 62, no. 6, pp. 1619-1628, 2009.
- [361] X. Zhang, X. Lu, Q. Shi, X.-q. Xu, E. L. Hon-chiu, L. N. Harris, J. D. Iglehart, A. Miron, J. S. Liu and W. H. Wong, "Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data," *BMC bioinformatics*, vol. 7, no. 1, 2006.
- [362] R. B. Bendel and A. A. Afifi, "Comparison of stopping rules in forward "stepwise" regression," *Journal of the American Statistical Association*, vol. 72, no. 357, pp. 46-53, 1977.
- [363] A. J. Miller, "Selection of subsets of regression variables," *Journal of the Royal Statistical Society. Series A (General)*, vol. 147, no. 3, pp. 389-425, 1984.
- [364] J. B. Wendt, F. Koushanfar and M. Potkonjak, "Techniques for Foundry Identification," *Proceedings of the 51st Annual Design Automation*

Conference , pp. 1-6, 2014.

- [365] A. Cord, C. Ambroise and J.-P. Cocquerez, "Feature selection in robust clustering based on Laplace mixture," *Pattern Recognition Letters*, vol. 27, no. 6, pp. 627-635, 2006.
- [366] P. Pitak-Arnnp, K. Dhanuthai and A. Hemprich, "Misleading p-value: Do you recognize it?," *European Journal of Dentistry*, vol. 4, pp. 356-358, 2010.
- [367] S. Senn, "P-values," *Encyclopedia of Biopharmaceutical Statistics*, pp. 685-695, 2003.
- [368] M. Pandolfi and G. Carreras, "The faulty statistics of complementary alternative medicine (CAM)," *European Journal of Internal Medicine*, vol. 25, pp. 607-609, 2014.
- [369] L. G. Halsey, D. Curran-Everett, S. L. Vowler and G. B. Drummond, "The fickle P value generates irreproducible results," *Nature Methods*, vol. 12, no. 3, pp. 179-185, 2015.
- [370] B. W. Ramsey, B. E. Mullins, R. Speers and K. A. Batterton, "Watching for Weakness in Wild WPANs," *Military Communications Conference (MILCOM)*, pp. 1404-1409, 2013.
- [371] J. N. Matthews and D. G. Altman, "Interaction 2: compare effect sizes not P values," *British Medical Journal (BMJ)*, vol. 313, p. 808, 28 Sept. 1996.
- [372] A. A. Neath, "Statistical inference, statistics education, and the fallacy of the transposed conditional," *ASA Section on Statistical Education*, pp. 3348-3350, 2010.
- [373] D. D. Boos and L. A. Stefanski, "P-value precision and reproducibility," *The American Statistician*, vol. 65, no. 4, pp. 213-221, 2011.
- [374] D. R. Anderson, W. A. Link, D. H. Johnson and K. P. Burnham, "Suggestions for presenting the results of data analysis," *The Journal of Wildlife Management*, vol. 65, no. 3, pp. 373-378, 2001.
- [375] K. Kitbumrungrat, "Comparison logistic regression and discriminant analysis in classification groups for breast cancer," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 12, no. 5, pp. 111-115, 2012.

- [376] J. Elith, J. R. Leathwick and T. Hastie, "A working guide to boosted regression trees," *Journal of Animal Ecology*, vol. 77, no. 4, pp. 802-813, 2008.
- [377] D. Trafimow and M. Marks, "Editorial," *Basic and Applied Sociate Psychology*, vol. 37, no. 1, pp. 1-2, 2015.
- [378] J. P. Simmons, L. D. Nelson and U. Simonsohn, "False-Positive Psychology : Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psychological Science*, pp. 1-8, 2011.
- [379] J. L. Moran and P. J. Solomon, "A farewell to p-values?," *Critical Care and Resuscitation*, vol. 6, pp. 130-137, 2004.
- [380] D. Morrison and R. Henkel, *The significance test controversy - a reader*, Chicago, IL: Aldine Publishing, 1969.
- [381] L. Harlow, S. Mulaik and J. Steiger, *What if there were no significance tests*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1997.
- [382] R. Nickerson, "Null hypothesis significance testing: a review of an old and continuing controversy," *Psychology Methods*, vol. 5, pp. 241-301, 2000.
- [383] D. Johnson, "The insignificance of statistical significance testing," *Journal of Wildlife Management*, vol. 63, pp. 763-772, 1999.
- [384] F. Fidler, N. Thomason, G. Cumming, S. Finch and J. Leeman, "Editors can lead researchers to confidence intervals, but they can't make them think," *Psychological Science*, vol. 15, no. 2, pp. 119-126, 2004.
- [385] R. Fisher, in *Statistical Methods for Research Workers*, 11th ed., Edinburgh, Oliver and Boyd, 1950, pp. 99-101.
- [386] S. Goodman, "Commentary: The p-value, devalued," *International Journal of Epidemiology*, vol. 32, pp. 699-702, 2003.
- [387] D. R. Anderson, K. P. Burnham and W. L. Thompson, "Null hypothesis testing: problems, prevalence, and an alternative," *The journal of wildlife management* , vol. 64, no. 4, pp. 912-923, 2000.
- [388] S. N. Goodman, "Toward evidence-based medical statistics. 1: The P value fallacy," *Annals of internal medicine*, vol. 130, no. 12, pp. 995-1004, 1999.

- [389] C. Y. Huak, "Are you a p-value worshipper?," *European journal of dentistry*, vol. 3, no. 3, 2009.
- [390] R. L. Koretz, "Is statistical significance always significant?," *Nutrition in clinical practice*, vol. 20, no. 3, pp. 303-307, 2005.
- [391] R. K.J., "Writing for epidemiology," *Epidemiology*, vol. 9, pp. 333-337, 1998.
- [392] G. L. Grunckenmeier, Y. Wu and A. P. Furnary, "What is the value of a p value?," *The Annals of Thoracic Surgery*, vol. 87, pp. 1337-1343, 2009.
- [393] S. Senn, "Two cheers for p-values?," *Journal of Epidemiology and Biostatistics*, vol. 6, no. 2, pp. 193-204, 2001.
- [394] V. Vovk, "Combining p-values via averaging," *arXiv preprint*, 2012.
- [395] X. Sheng and J. Yang, "An adaptive truncated product method for combining dependent p-values," *Economics Letters*, vol. 119, pp. 180-182, 2013.
- [396] M. B. Brown, "A Method for Combining Non-Independent, One-Sided Tests of Significance," *Biometrics*, vol. 31, no. 4, pp. 987-992, 1975.
- [397] T. M. Loughin, "A systematic comparison of methods for combining p-values from independent tests," *Computational Statistics & Data Analysis*, vol. 47, no. 3, pp. 467-485, 2004.
- [398] D. V. Zaykin, L. A. Zhivotovsky, P. H. Westfall and B. S. Weir, "Truncated product method for combining P-values," *Genetic epidemiology*, vol. 22, no. 2, pp. 170-185, 2002.
- [399] R. C. Elston, "On Fisher's Method of Combining p-Values," *Biometrical journal*, vol. 33, no. 3, pp. 339-345, 1991.
- [400] E. S. Edgington, "An Additive Method for Combining Probability Values from Independent Experiments," *The Journal of Psychology: Interdisciplinary and Applied*, vol. 80, no. 2, pp. 351-363, 1972.
- [401] R. D. Cousins, "Annotated bibliography of some papers on combining significances or p-values," *arXiv preprint*, vol. 0705.2209, 2009.
- [402] S. Won, N. Morris, Q. Lu and R. C. Elston, "Choosing an Optimal Method to Combine P-values," *Statistics in Medicine*, vol. 28, no. 11, p. 1537-1553,

2009.

- [403] A. M. Winkler, S. M. Smith and O. E. Nichols, "Non-Parametric Combination for Analyses of Multi-Modal Imaging," *Neuroimaging Data Analysis. Statistical and Applied Mathematical Sciences Institute (SAMSI)*, 4-14 Jun 2013.
- [404] E. Martz, "Three Things the P-Value Can't Tell You about Your Hypothesis Test," 20 June 2011. [Online]. Available: <http://blog.minitab.com/blog/understanding-statistics/three-things-the-p-value-cant-tell-you-about-your-hypothesis-test>. [Accessed 10 Feb. 2015].
- [405] A. Gelman, "Simulation of a statistics blogosphere," 22 Mar. 2010. [Online]. Available: http://andrewgelman.com/2010/03/22/simulation_of_a/. [Accessed 10 Feb. 2015].
- [406] J. O. Berger and T. Sellke, "Testing a point null hypothesis: the irreconcilability of P values and evidence," *Journal of the American statistical Association*, vol. 82, no. 397, pp. 112-122, 1987.
- [407] M. H. Kutner, C. J. Nachtsheim, J. Neter and W. Li, "Normal Error Regression Model," in *Applied Linear Statistical Models*, McGraw Hill Irwin, 2005, pp. 26-32.
- [408] C. Moler, "Floating Points," *MATLAB News and Notes*, pp. 1-3, 1996.
- [409] R. I. Jennrich, "Stepwise Discriminant Analysis," in *Statistical Methods for Digital Computers*, New York, Wiley, 1977, pp. 76-95.
- [410] R. J. Calantone and R. G. Cooper, "A discriminant model for identifying scenarios of industrial new product failure," *Journal of the Academy of Marketing Science*, vol. 7, no. 3, pp. 163-183, 1979.
- [411] M. Kupinski and M. L. Giger, "Feature selection and classifiers for the computerized detection of mass lesions in digital mammography," *International Conference on Neural Networks*, vol. 4, pp. 2460-2463, 1997.
- [412] G. Castaldo, G. Oriani, L. Cimino, M. Topa, G. Budillon, F. Salvatore and L. Sacchetti, "Discriminant function based on serum analytes differentiates hepatocarcinoma from secondary liver neoplasia," *Clinical chemistry*, vol. 43, no. 3, pp. 439-443, 1995.

- [413] J. M. Santoro, Identifying Innovativeness Among Users of Wireless Features and Services, North Carolina State University: MS Thesis, 2001.
- [414] M. Domínguez-Rodrigo, H. T. Bunn and J. Yravedra, "A critical re-evaluation of bone surface modification models for inferring fossil hominin and carnivore interactions through a multivariate approach: application to the FLK Zinj archaeofaunal assemblage (Olduvai Gorge, Tanzania)," *Quaternary International*, vol. 322, pp. 32-43, 2014.
- [415] A. O. Finley, D. B. Kittredge, T. H. Stevens, C. M. Schweik and D. C. Dennis, "Interest in cross-boundary cooperation: Identification of distinct types of private forest owners," *Forest Sci.*, vol. 52, no. 1, pp. 10-22, 2006.
- [416] C. J. Watson, "Approaches for the interpretation of multiple discriminant analysis in organizational research," *Academy of Management Review*, vol. 7, no. 1, pp. 124-130, 1982.
- [417] T. J. Bihl, K. W. Bauer and M. A. Temple, "A Comparison of Feature Selection Methods for RF-DNA Fingerprinting Using ZigBee Device Emissions," *IEEE Transactions on Information Forensics and Security*, submitted: July, 2015.
- [418] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6-23, 1997.
- [419] M. B. Crawford, Shop Class as Soulcraft: An Inquiry Into the Value of Work, Penguin Press, 2009.
- [420] Aristophanes and M. W. Humphreys, Aristophanes: Clouds, Ginn, Heath, & Company, 1892.
- [421] B. R. Turnquist, Fusion Schemes for Ensembles of Hyperspectral Anomaly Detection Algorithms, Air Force Institute of Technology: MS Thesis, 2011.
- [422] X. Zhao, X. Li, Z. Ma and M. Yin, "Prediction of Lysine Ubiquitylation with Ensemble Classifier and Feature Selection," *International Journal of Molecular Science*, vol. 12, pp. 8347-8361, 2011.
- [423] H. K. Butler, The Relationship Between Diversity and Accuracy in Multiple Classifier Systems, Air Force Institute of Technology: MS Thesis, 2012.

- [424] W. D. Perreault, D. N. Behrman and G. M. Armstrong, "Alternative Approaches for Interpretation of Multiple Discriminant Analysis in Marketing Research," *Journal of Business Research*, vol. 7, pp. 151-173, 1979.
- [425] S. K. Darden, T. Dabelsteen and S. B. Pedersen, "A potential tool for swift fox (*Vulpes velox*) conservation: individuality of long-range barking sequences," *Journal of Mammalogy*, vol. 84, no. 4, pp. 1417-1427, 2003.
- [426] J. C. Hosseini and R. L. Armacost, "Marketing Strategy Implications of Consumer Preferences For Downtown Housing," *Academy of Marketing Science (AMS) Annual Conference*, pp. 360-364, 1989.
- [427] T. Ramayah, J. Ignatius, J. Y. A. Leen and L. M. Chiun, "Discriminant Analysis," in *Probability and Statistics a Didactic Introduction*, CRC Press, 2014, pp. 384-415.
- [428] T. Gruca, D. Nath and H. Thomas, "Identifying and Comparing Strategic Groups Using Alternative Methods: Method Validation and Group Convergence in a Single Mature Industry," in *Statistical Models for Strategic Management*, Kluwer Academic Publishers, 1997, pp. 55-80.
- [429] W. D. Parker, "An Empirical Typology of Perfectionism in Academically Talented Children," *American Educational Research Journal*, vol. 34, no. 3, pp. 545-562, 1997.
- [430] W. D. Parker, "Healthy perfectionism in the gifted," *Prufrock Journal*, vol. 11, no. 4, pp. 173-182, 2000.
- [431] M. E. Ramsey, T. L. Maginnis, R. Y. Wong, C. Brock and M. E. Cummings, "Identifying context-specific gene profiles of social, reproductive, and mate preference behavior in a fish species with female mate choice," *Frontiers in neuroscience*, vol. 6, pp. 1-15, 2012.
- [432] J. P. Cannon and W. D. Perreault, "Buyer-Seller Relationships in Business Markets," *Journal of Marketing Research*, vol. 36, no. 4, pp. 439-460, 1999.
- [433] S. B. Kim and P. Rattakorn, "Unsupervised feature selection using weighted principal components," *Expert Systems with Applicat.*, vol. 38, no. 5, pp. 5704-5710, 2011.
- [434] A. Jain, K. Nandakumar and A. Ross, "Score normalization in multimodal

- biometric systems," *Pattern Recognition*, vol. 38, pp. 2270-2285, 2005.
- [435] J. M. Haning, Feature selection for high-dimensional individual and ensemble classifiers with limited data, University of Cincinnati: MS Thesis, 2014.
- [436] F. Ortiz, "Dealing with Categorical Data Types in a Designed Experiment Part I: Why You Should Avoiding Using Categorical Data Types," STAT T&E Center of Excellence, Wright Patterson AFB, OH, 2012.
- [437] F. Ortiz, "Dealing with Categorical Data Types in a Designed Experiment Part II: Sizing a Designed Experiment When Using a Binary Response," STAT T&E Center of Excellence, Wright Patterson AFB, OH, 2014.
- [438] V. Fedorov, F. Mannino and R. Zhang, "Consequences of dichotomization," *Pharmaceutical Statistics*, vol. 8, pp. 50-61, 2009.
- [439] J. Cohen, "The Cost of Dichotomization," *Applied Psychological Measurement*, vol. 7, no. 3, pp. 249-253, 1983.
- [440] M. Hamada, "The Advantages of Continuous Measurements over Pass/Fail Data," *Quality Engineering*, vol. 15, no. 2, pp. 253-258, 2002.
- [441] P. Royston, D. G. Altman and W. Sauerbrei, "Dichotomizing continuous predictors in multiple regression: a bad idea," *Statistics in Medicine*, vol. 25, no. 1, pp. 127-141, 2006.
- [442] C. Sanderson and K. K. Paliwal, "Joint cohort normalization in a multi-feature speaker verification system," *IEEE International Conference on Fuzzy Systems*, pp. 232-235, 2001.
- [443] C. Sanderson and K. K. Paliwal, "Information fusion for robust speaker verification," *INTERSPEECH*, pp. 755-758, 2001.
- [444] C. C. Toh, B. Zhang and Y. Wang, "Multiple-feature fusion based onset detection for solo singing voice," *ISMIR*, pp. 515-520, 2008.
- [445] A. Gyaourova and A. Ross, "A coding scheme for indexing multimodal biometric databases," *Computer Vision and Pattern Recognition Conference (CVPR)*, 2009.
- [446] S. Chen and Y. Tian, "Margin-constrained multiple kernel learning based multi-modal fusion for affect recognition," *IEEE International Conference and*

Workshops on Automatic Face and Gesture Recognition, pp. 1-7, 2013.

- [447] X. Zhou and B. Bhanu, "Feature fusion of face and gait for human recognition at a distance in video," *International Conference on Pattern Recognition*, pp. 529-532, 2006.
- [448] G. Feng, K. Dong, D. Hu and D. Zhang, "When faces are combined with palmprints: a novel biometric fusion strategy," *Biometric authentication*, pp. 701-707, 2004.
- [449] T. Kinnunen, V. Hautamäki and P. Fränti, "Fusion of spectral feature sets for accurate speaker identification," *International Conference on Speech and Computer*, pp. 361-365, 2004.
- [450] A. Kumar, D. C. Wong, H. C. Shen and A. K. Jain, "Personal verification using palmprint and hand geometry biometric," *Audio-and Video-Based Biometric Person Authentication*, pp. 668-678, 2003.
- [451] A. Ross and R. Govindarajan, "Feature level fusion of hand and face biometrics," *SPIE Conference on Biometric Technology for Human Identification II*, pp. 196-204, 2005.
- [452] J. Yang, J. Yang, D. Zhang and J. Lu, "Feature fusion: parallel strategy vs. serial strategy," *Pattern Recognition*, vol. 38, no. 6, pp. 1369-1381, 2003.
- [453] S. Chen, Y. Tian, Q. Liu and D. N. Metaxas, "Recognizing expressions from face and body gesture by temporal normalized motion and appearance features," *Image and Vision Computing*, vol. 31, no. 2, pp. 175-185, 2013.
- [454] H. Gunes and M. Piccardi, "Automatic temporal segment detection and affect recognition from face and body display," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 1, pp. 64-84, 2009.
- [455] C. Shan, S. Gong and P. W. McOwan, "Beyond Facial Expressions: Learning Human Emotion from Body Gestures," *British Machine Vision Conference (BMVC)*, pp. 1-10, 2007.
- [456] L. Gupta, B. Chung, M. D. Srinath, D. L. Molfese and H. Kook, "Multichannel fusion models for the parametric classification of differential brain activity," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 11, pp. 1869-1881, 2005.

- [457] H. Kekre, T. K. Sarode and J. K. Save, "Effect of distance measures on image transform based image classification," *International Journal of Engineering Science and Technology*, vol. 4, no. 8, pp. 3729-3742, 2012.
- [458] D. A. Jackson, "Stopping rules in principal component analysis: a comparison of heuristical and statistical approaches," *Ecology*, vol. 74, no. 8, pp. 2204-2214, 1993.
- [459] K. Bauer, A Monte Carlo study of dimensionality assessment and factor interpretation in principal component analysis, Air Force Institute of Technology: MS Thesis, 1981.
- [460] P. R. Peres-Neto, D. A. Jackson and K. M. Somers, "How many principal components? Stopping rules for determining the number of non-trivial axes revisited," *Computational Statistics & Data Analysis*, vol. 49, no. 4, pp. 974-997, 2005.
- [461] W. R. Zwick and W. F. Velicer, "Comparison of five rules for determining the number of components to retain," *Psychological bulletin*, vol. 99, no. 3, 1986.
- [462] G. R. Foxall and G. E. Greenley, "Predicting and explaining responses to consumer environments: An empirical test and theoretical extension of the behavioural perspective model," *The Service Ind. J.*, vol. 20, no. 2, pp. 39-63, 2000.
- [463] J. Hair, R. Anderson, R. Tathan and W. Black, *Multivariate Data Analysis*, London: Prentice-Hall, 1995.
- [464] W. Klecka, *Discriminant Analysis*, Newbury Park, CA: Sage, 1980.
- [465] M. Cowles and C. Davis, "On the Origins of the .05 Level of Statistical Significance," *American Psychologist*, vol. 35, no. 5, pp. 553-558, 1982.
- [466] A. L. Bigley, *Horn's Curve Estimation Through Multi-Dimensional Interpolation*, Air Force Institute of Technology: MS Thesis, 2013.
- [467] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433-459, 2010.

- [468] R. J. Johnson, J. P. Williams and K. W. Bauer, "AutoGAD: An improved ICA-based hyperspectral anomaly detection algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 6, pp. 3492-3503, 2013.
- [469] J. L. Horn, "A rationale and test for the number of factors in factor analysis," *Psychometrika*, vol. 30, no. 2, pp. 179-185, 1965.
- [470] M. Biehl, B. Hammer, P. Schneider and T. Villmann, "Metric Learning for Prototype-Based Classification," in *Innovations in Neural Information Paradigms and Applications*, Springer, 2009.
- [471] J. V. M. Lee, *Nonlinear dimensionality reduction*, Heidelberg: Springer, 2007.
- [472] Z. Chuan, L. Xianliang, H. Mengshu and Z. Xu, "A LVQ-based neural network anti-spam email approach," *ACM SIGOPS Operating Systems Review*, pp. 34-39, 2005.
- [473] Z. Chuan, L. Xianliang and X. Qian, "A Novel Anti-spam Email Approach Based on LVQ," *Parallel and Distributed Computing: Applications and Technologies*, pp. 180-183, 2005.
- [474] H. S. Behera, D. K. Acharya and S. S. Panda, "Modified Linear Vector Quantization Technique for Classification of Heart Disease Data," *International Journal of Advanced Research in Computer Science*, vol. 3, no. 4, pp. 315-319, 2012.
- [475] Z. Lu and T. Peng, "The VoIP intrusion detection through a LVQ-based neural network," *IEEE International Conference for Internet Technology and Secured Transactions (ICITST)*, pp. 1-6, 2009.
- [476] J. Ji, D. Kunita and Q. Zhao, "A Customer Intention Aware System for Document Analysis," *International Joint Conference on Neural Networks (IJCNN)*, pp. 1-6, 2010.
- [477] J. Ji, D. Kunita and Q. Zhao, "A human behavior inspired awareness system for document analysis," *IEEE International Symposium on Aware Computing (ISAC)*, pp. 1-6, 2010.
- [478] B. Widrow and M. Hoff, "Adaptive Switching Circuits," *Western Electric Show and Convention Record*, pp. 96-104, 1960.

- [479] J. K. Hunter, "Differentiable Functions," in *An Introduction to Real Analysis*, 2012, pp. 39-55.
- [480] X. Zhang, C.-C. Hang, S. Tan and P.-Z. Wang, "The min-max function differentiation and training of fuzzy neural networks," *IEEE Transactions on Neural Networks*, vol. 7, no. 5, pp. 1139-1150, 1996.
- [481] F.-M. Schleif, T. Villmann, B. Hammer, P. Schneider and M. Biehl, "Generalized Derivative Based Kernelized Learning Vector Quantization," *IDEAL*, pp. 21-28, 2010.
- [482] G. W. Milligan and M. C. Cooper, "Methodology review: Clustering methods," *Applied psychological measurement*, vol. 11, no. 4, pp. 329-354, 1987.
- [483] M. Forina, C. Armanino and V. Raggio, "Clustering with dendrograms on interpretation variables," *Analytica Chimica Acta*, vol. 454, no. 1, pp. 13-19, 2002.
- [484] J. F. Mello and M. A. Buzas, "An application of cluster analysis as a method of determining biofacies," *Journal of Paleontology*, pp. 747-758, 1968.
- [485] P. Sneath, "The application of computer to taxonomy," *Journal of General Microbiology*, vol. 17, pp. 201-226, 1957.
- [486] J. A. East, K. W. Bauer Jr and J. W. Lanning, ""Feature selection for predicting pilot mental workload: A feasibility study," *International Journal of Smart Engineering System Design*, vol. 4, no. 3, pp. 183-193, 2002.
- [487] R. Neches, R. E. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator and W. R. Swartout, "Enabling technology for knowledge sharing," *AI magazine*, vol. 12, no. 3, pp. 36-56, 1991.
- [488] F. Gazzaniga, R. Stebbins, S. Z. Chang, M. A. McPeck and C. Brenner, "Microbial NAD metabolism: lessons from comparative genomics," *Microbiology and molecular biology reviews*, vol. 73, no. 3, pp. 529-541, 2009.
- [489] D. R. Branson, D. N. Armentrout, W. M. Parker, C. van Hall and L. I. Bone, "Effluent monitoring step by step.," *Environmental science & technology*, vol. 15, no. 5, pp. 513-518, 1981.
- [490] J. P. Martino, "Looking ahead with confidence," *IEEE Spectrum*, pp. 76-81, Mar.

1985.

- [491] A. Farmer and P. McGuffin, "The classification of the depressions. Contemporary confusion," *The British Journal of Psychiatry*, vol. 155, pp. 437-443, 1989.
- [492] G. P. Styan, "Hadamard products and multivariate statistical analysis," *Linear algebra and its applications*, vol. 6, pp. 217-240, 1973.
- [493] J. W. Clark, "Neural networks: New tools for modeling and data analysis," in *Scientific Applications of Neural Nets*, Springer Berlin Heidelberg, 1999, pp. 1-96.
- [494] P. R. Gorman and T. Sejnowski, "Analysis of hidden units in a layered network trained to classify sonar targets," *Neural Networks*, pp. 75-89, 1988.
- [495] E. Patuwo, M. Y. Hu and M. S. Hung, "Two-Group Classification Using Neural Networks," *Decision Sciences*, vol. 24, no. 4, pp. 825-845, 1993.
- [496] D. Fletcher and E. Goss, "Forecasting With Neural Networks: An Application Using Bankruptcy Data," *Information and Management*, vol. 24, no. 3, pp. 159-167, 1993.
- [497] M. S. Hung, M. Y. Hu, M. S. Shanker and B. E. Patuwo, "Estimating Posterior Probabilities in Classification Problems With Neural Networks," *International Journal of Computational Intelligence and Organizations*, vol. 1, no. 1, pp. 49-60, 1996.
- [498] S. Chattopadhyay and G. Chattopadhyay, "Identification of the best hidden layer size for three-layered neural net in predicting monsoon rainfall in India," *Journal of Hydroinformatics*, vol. 10, no. 2, pp. 181-188, 2008.
- [499] J. C. Lindsey, A. M. Herzberg and D. G. Watts, "A method for cluster analysis based on projections and quantile-quantile plots," *Biometrics*, vol. 43, no. 2, pp. 327-341, 1987.
- [500] D. J. Hand, F. Daly, K. McConway, D. Lunn and E. Ostrowski, *A Handbook of Small Data Sets*, CRC Press, 1993.
- [501] P. Gao, C. Chen and S. Qin, "An Optimization Method of Hidden Nodes for Neural Network," *International Workshop on Education Technology and Computer Science*, pp. 53-56, 2010.

- [502] J. Mao, K. Mohiuddin and A. Jain, "Parsimonious Network Design and Feature Selection Through Node Pruning," pp. 622-624, 1994.
- [503] L. Zhang, J.-H. Jiang, P. Liu, Y.-Z. Liang and R.-Q. Yu, "Multivariate nonlinear modelling of fluorescence data by neural network with hidden node pruning algorithm," *Analytic Chimica Acta*, pp. 29-39, 1997.
- [504] R. Reed, "Pruning algorithms-a survey," *IEEE Transactions on Neural Networks*, vol. 4, no. 5, pp. 740-747, 1993.
- [505] S. Walczak and N. Cerpa, "Heuristic principals for the design of artificial neural networks," *Information and Software Technology*, vol. 41, no. 2, pp. 109-119, 1999.
- [506] B. A. Jain and B. R. Nag, "Artificial Neural Network Models for Pricing Initial Public Offerings," *Decision Sciences*, vol. 26, no. 3, pp. 283-302, 1995.
- [507] M. J. Lenard, P. Alam and G. R. Madey, "The Application of Neural Networks and a Qualitative Response Model to the Auditor's Going Concern Uncertainty Decision," *Decision Sciences*, vol. 26, no. 2, pp. 209-227, 1995.
- [508] S. Piramuthu, M. Shaw and J. Gentry, "Classification approach using multi-layered neural networks," *Decision Support Systems*, vol. 11, no. 5, pp. 509-525, 1994.
- [509] G. Mirchandani and W. Cao, "On hidden nodes for neural nets," *IEEE Transactions on Circuits and Systems*, vol. 36, no. 5, pp. 661-664, 1989.
- [510] M. Hayashi, "A fast algorithm for the hidden units in a multilayer perceptron," *Proc. IEEE 1993 Int. Joint Conf. Neural Networks*, pp. 351-354, 1993.
- [511] E. Patuwo, M. Y. Hu and M. S. Hung, "Two-Group Classification Using Neural Networks," *Decision Sciences*, vol. 24, no. 4, pp. 825-845, 1993.
- [512] G. Daqi and W. Shouyi, "An optimization method for the topological structures of feed-forward multi-layer neural networks," *Pattern Recognition*, pp. 1337-1342, 1998.
- [513] T. J. Paciencia, Improving Non-Linear Approaches to Anomaly Detection, Class Separation, and Visualization, Air Force Institute of Technology: PhD Dissertation, 2014.

- [514] J. P. Williams, Towards the Mitigation of Correlation Effects in the Analysis of Hyperspectral Imagery with Extensions to Robust Parameter Design, Air Force Institute of Technology: PhD Dissertation, 2012.
- [515] J. P. Bellucci, T. E. Smetek and K. W. Bauer, "Improved hyperspectral image processing algorithm testing using synthetic imagery and factorial designed experiments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 3, pp. 1211-1223, 2010.
- [516] J. P. Bellucci, Improved hyperspectral image testing using synthetic imagery and factorial designed experiments., Air Force Institute of Technology: MS Thesis, 2007.
- [517] F. M. Mindrup, Optimizing Hyperspectral Imagery Anomaly Detection through Robust Parameter Design, Air Force Institute of Technology: PhD Dissertation, 2011.
- [518] F. M. Mindrup, T. J. Bihl and K. W. Bauer, "Modeling noise in a framework to optimize the detection of anomalies in hyperspectral imaging," *Intelligent Engineering Systems through Artificial Neural Networks: Computational Intelligence in Architecting Complex Engineering Systems*, vol. 20, pp. 517-524, 2010.
- [519] C.-C. Chiu, D. F. Cook, J. P. Pignatiello and A. D. Whittaker, "Design of a radial basis function neural network with a radius-modification algorithm using response surface methodology," *Journal of Intelligent Manufacturing*, vol. 8, no. 2, pp. 117-124, 1997.
- [520] C.-C. Chiu and C.-T. Su, "A Novel Neural Network Model Using Box-Jenkins Technique and Response Surface Methodology to Predict Unemployment Rate," *IEEE International Conference on Tools with Artificial Intelligence*, pp. 74-80, 1998.
- [521] R. Byrd, J. C. Gilbert and J. Nocedal, "A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming," *Mathematical Programming*, vol. 89, no. 1, pp. 149-185, 2000.
- [522] R. Byrd, M. E. Hribar and J. Nocedal, "An Interior Point Algorithm for Large-Scale Nonlinear Programming," *SIAM Journal on Optimization*, vol. 9, no. 4, pp. 877-900, 1999.

- [523] R. A. Waltz, J. L. Morales, J. Nocedal and D. Orban, "An interior algorithm for nonlinear optimization that combines line search and trust region steps," *Mathematical Programming*, vol. 107, no. 3, pp. 391-408, 2006.
- [524] M. Powell, "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations," in *Numerical Analysis, Lecture Notes in Mathematics*, Springer Verlag,, 1978.
- [525] M. Powell, "The Convergence of Variable Metric Methods For Nonlinearly Constrained Optimization Calculations," in *Nonlinear Programming 3*, Academic Press, 1978.
- [526] S. Han, "A Globally Convergent Method for Nonlinear Programming," *Journal of Optimization Theory and Applications*, vol. 22, no. 3, pp. 297-309, 1977.
- [527] B. Picinbono, "On instantaneous amplitude and phase of signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 552-560, 1997.
- [528] R. L. Dowden and C. D. D. Adams, "Size and location of lightning-induced ionisation enhancements from measurement of VLF phase and amplitude perturbations on multiple antennas," *Journal of atmospheric and terrestrial physics*, vol. 55, no. 10, pp. 1335-1359, 1993.
- [529] E. H. Armstrong, "A method of reducing disturbances in radio signaling by a system of frequency modulation," *Proceedings of the Institute of Radio Engineers*, vol. 24, no. 5, pp. 689-740, 1936.
- [530] A. Hajimiri and T. H. Lee, "A general theory of phase noise in electrical oscillators," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 2, pp. 179-194, 1998.
- [531] D. F. Williamson, R. A. Parker and J. S. Kendrick, "The box plot: a simple visual method to interpret data," *Annals of internal medicine*, vol. 110, no. 11, pp. 916-921, 1989.
- [532] R. McGill, J. W. Tukey and W. A. Larsen, "Variations of Boxplots," *The American Statistician*, vol. 32, no. 1, pp. 12-16, 1978.
- [533] W. L. Martinez, A. Martinez and J. Solka, *Exploratory data analysis with MATLAB*, CRC Press, 2010.

- [534] A. Jain, K. Nandakumar and A. Ross, "Score normalization in multimodal biometric systems," *Pattern recognition*, vol. 38, no. 12, pp. 2270-2285, 2005.
- [535] A. Mendez, A. Belghith and M. Sawan, "A DSP for sensing the bladder volume through afferent neural pathways," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, no. 4, pp. 552-564, 2014.
- [536] C. C. Olson, Evolutionary algorithms, chaotic excitations, and structural health monitor: on global search methods for improved damage detection via tailored inputs, PhD Dissertation: UC San Diego, 2008.
- [537] M. Freed, J. A. deZwart, P. Hariharan, M. R. Myers and A. Badano, "Development and characterization of a dynamic lesion phantom for the quantitative evaluation of dynamic contrast-enhanced MRI," *Medical Physics*, vol. 38, no. 10, pp. 5601-5611, 2011.
- [538] K. Kipp, R. Pfeiffer, M. Sabick, C. Harris, J. Sutter, S. Kuhlman and K. Shea, "Muscle Synergies During a Single-Leg Drop-Landing in Boys and Girls," *Journal of Applied Biomechanics*, vol. 30, pp. 262-268, 2014.
- [539] D. Colombi, B. Thors, N. Wiren, L.-E. Larsson and C. Törnevik, "Measurements of downlink power level distributions in LTE networks," *International Conference on Electromagnetics in Advanced Applications (ICEAA)*, pp. 98-101, 2013.
- [540] R. L. A. Batista, M. M. Franco, L. M. V. Naldoni, G. Duarte, A. S. Oliveira and C. H. J. Ferreira, "Biofeedback and the electromyographic activity of pelvic floor muscles in pregnant women," *Brazilian Journal of Physical Therapy*, vol. 15, no. 5, pp. 386-392, 2011.
- [541] S. Akselrod, V. Mor-Avi and D. David, "Method of measuring regional tissue blood flow". US Patent 5135000, 1991.
- [542] J. Lawrence, "Data preparation for a neural network," *AI Expert: Neural Network Special Report*, vol. 6, no. 11, pp. 15-21, 1991.
- [543] R. Snelick, U. Uludag, A. Mink, M. Indovina and A. Jain, "Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 450-455, 2005.

- [544] D. P. Bischak, "Weighted batch means and improvements in coverage," *Winter Simulation Conference*, pp. 474-480, 1993.
- [545] B. L. Fox, D. Goldsman and J. J. Swain, "Spaced batch means," *Operations Research Letters*, vol. 10, pp. 255-263, 1991.
- [546] E. K. Lada, N. M. Steiger and J. R. Wilson, "SBatch: A spaced batch means procedure for steady-state simulation analysis," *Journal of Simulation*, vol. 2, pp. 170-185, 2008.
- [547] E. K. Lada and J. R. Wilson, "SBatch: A spaced batch means process for simulation analysis," *Winter Simulation Conference*, pp. 463-471, 2007.
- [548] F. Yee, "A Study on "Gap-means," Batch Means," and "Weighted Batches"," OPER 760 - Analysis Project Report, Air Force Institute of Technology, 2010.
- [549] S. R. Das, D. Duffie, N. Kapadia and L. Saita, "Common failings: How corporate defaults are correlated," *The Journal of Finance*, vol. 62, no. 1, pp. 93-117, 2007.
- [550] A. Tafazzoli and J. R. Wilson, "Skart: A skewness- and autoregression-adjusted batch-means procedure for simulation analysis," *IIE Transactions*, vol. 43, pp. 110-128, 2011.
- [551] P. D. Welch, "The statistical analysis of simulation results.," in *The computer performance modeling handbook* , 1983, pp. 268-328..
- [552] M. Sherman, "On databased choice of batch size for simulation output analysis," *Simulation*, vol. 71, no. 1, pp. 38-47, 1998.
- [553] C. Chien, "Batch size selection for the batch means method," *Winter Simulation Conference*, pp. 345-352, 1994.
- [554] B. Schmeiser, "Batch size effects in the analysis of simulation output," *Operations Research*, vol. 30, no. 3, pp. 556-568, 1982.
- [555] G. A. Hawawini, "An analytical examination of the intervaling effect on skewness and other moments," *Journal of Financial and Quantitative Analysis*, vol. 15, no. 5, pp. 1121-1127, 1980.

- [556] N. M. Steiger and J. R. Wilson, "An improved batch means procedure for simulation output analysis," *Management Science*, vol. 48, no. 12, pp. 1569-1586, 2002.
- [557] W. T. Song and M. Chih, "Run length not required: Optimal-mse dynamic batch means estimators for steady-state simulations," *European Journal of Operations Research*, vol. 229, pp. 114-123, 2013.
- [558] W. T. Song, "On the estimation of optimal batch sizes in the analysis of simulation output," *European Journal of Operational Research*, vol. 88, pp. 304-319, 1996.
- [559] J. Dougherty, R. Kohavi and M. Sahami, "Supervised and unsupervised discretization of continuous features," *ICML*, pp. 194-202, 1995.
- [560] P. E. Anderson, D. A. Mahle, T. E. Doom, N. V. Reo, N. J. DelRaso and M. L. Raymer, "Dynamic adaptive binning: an improved quantification technique for NMR spectroscopic data," *Metabolomics*, vol. 7, no. 2, pp. 179-190, 2011.
- [561] I. Sibomana, 'Functional Metabolomics' Enhances Assessment of Tissue Dysfunction as Demonstrated in a Rate Model of Sub-Acute D-Serine Exposure, MS Thesis: Wright State University, 2011.
- [562] U. Vishkin, "Deterministic Sampling - A New Technique for Fast Pattern Matching," *SIAM Journal on Computing*, vol. 20, no. 1, pp. 22-40, 1991.
- [563] D. Roobaert, G. Karakoulas and N. V. Chawla, "Information Gain, Correlation and Support Vector Machines," in *Feature Extraction*, Springer Berlin Heidelberg, 2006, pp. 463-470.
- [564] L. Tološi and T. Lengauer, "Classification with correlated features: unreliability of feature ranking and solutions," *Bioinformatics*, vol. 27, no. 14, pp. 1986-1994, 2011.
- [565] W. C. Smith and B. Manhire, "Statistical and frequency domain analysis of power system loads," *North American Power Symposium*, 1992.
- [566] J. Biesiada and W. Duch, "Feature Selection for High-Dimensional Data: A Pearson Redundancy Based Filter," *Computer Recognition Systems*, vol. 2, pp. 242-249, 2007.

- [567] L. Yu and H. Liu, "Efficiently Handling Feature Redundancy in High Dimensional Data," *SIGKDD*, pp. 685-690, 2003.
- [568] D. G. Morrison, "On the interpretation of discriminant analysis," *Journal of Marketing Research*, vol. 6, pp. 156-163, 1969.
- [569] M. Feindt and U. Kerzel, "The NeuroBayes neural network package," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 559, no. 1, pp. 190-194, 2006.
- [570] G. S. d. S. Gomes, T. B. Ludermir and L. M. M. R. Lima, "Comparison of new activation functions in neural network for forecasting financial time series," *Neural Computing and Applications*, vol. 20, no. 3, pp. 417-439, 2011.
- [571] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153-158, 1997.
- [572] G. Hinton, "Reducing the Dimensionality of Data with Neural Networks," *Science*, pp. 504-507, July 2007.
- [573] J. V. Stone, *Independent Component Analysis*, Cambridge, MA: The MIT Press, 2004.
- [574] T.-K. Yang, W.-M. Song and M.-C. Chih, "Revisit the dynamic-batch-means estimator," *Journal of the Chinese Institute of Industrial Engineers*, vol. 26, no. 6, pp. 499-509, 2009.
- [575] E. Mota, A. Wolisz and K. Pawlikowski, "A perspective of batching methods in a simulation environment of multiple replications in parallel," *Winter Simulation Conference*, pp. 761-766, 2000.
- [576] B. Arnrich, *Data Mart Based Research in Heart Surgery*, PhD Dissertation: Universität Bielefeld, 2006.
- [577] F. Yee, "A Study on "Gap-means," Batch Means," and "Weighted Batches"," *OPER 760 - Analysis Project Report*, Air Force Institute of Technology, Wright Patterson AFB, OH, 2010.
- [578] M. S. Meketon and B. Schmeiser, "Overlapping batch means: Something for

- nothing?," *Winter Simulation Conference*, pp. 226-230, 1984.
- [579] T. J. Paciencia, T. J. Bihl and K. W. Bauer, "Improved n-Dimensional Data Visualization from Hyper-Radial Values," *Journal of Computational and Graphical Statistics*, submitted 2015.
- [580] J. C. Bezdek, J. M. Keller, R. Krishnapuram, L. I. Kuncheva and N. R. Pal, "Will the real iris data please stand up?," *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 3, pp. 368-369, 1999.
- [581] A. R. Rocha Neto and G. A. Barreto, "On the Application of Ensembles of Classifiers to the Diagnosis of Pathologies of the Vertebral Column: A Comparative Analysis," *IEEE Latin America Transactions*, vol. 7, no. 4, pp. 487-496, 2009.
- [582] P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis, "Modeling wine preferences by data mining from physicochemical properties," *Decision Support Systems*, vol. 47, no. 4, pp. 547-553, 2009.
- [583] W. Wolberg and O. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," *Proceedings of the National Academy of Sciences*, pp. 9193-9196, 1990.
- [584] K. Bache and M. Lichman, "UCI Machine Learning Repository," 2013. [Online]. Available: <http://archive.ics.uci.edu/ml>.
- [585] Y. LeCun, C. Cortes and C. J. Burges, "The MNIST Database of Handwritten Digits," [Online]. Available: <http://yann.lecun.com/exdb/mnist/>. [Accessed 10 Dec. 2014].
- [586] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
- [587] P. Horton and K. Nakai, "A Probablistic Classification System for Predicting the Cellular Localization Sites of Proteins," *Intelligent Systems in Molecular Biology*, pp. 109-115, 1996.
- [588] P. Berry, *Roman Handwriting at the Time of Christ*, Edwin Mellen Press, 2001.
- [589] G. Mori and J. Malik, "Recognizing objects in adversarial clutter: Breaking a visual

CAPTCHA," *Computer Vision and Pattern Recognition*, 2003.

- [590] L. Von Ahn, B. Maurer, C. McMillen, D. Abraham and M. Blum, "recaptcha: Human-based character recognition via web security measures," *Science*, vol. 321, no. 5895, pp. 1465-1468, 2008.
- [591] Y. Leydier, F. Lebourgeois and H. Emptoz, "Text search for medieval manuscript images," *Pattern Recognition*, vol. 40, no. 12, pp. 3552-3567, 2007.
- [592] A. Fischer, M. Wuthrich, M. Liwicki, V. Frinken, H. Bunke, G. Viehhauser and M. Stolz, "Automatic transcription of handwritten medieval documents," *International Conference on Virtual Systems and Multimedia (VSMM)*, pp. 137-142, 2009.
- [593] SAS Institute, JMP User Guide, Cary, NC, 2009.
- [594] W. P. Jones and G. W. Furnas, "Pictures of relevance: A geometric analysis of similarity measures," *Journal of the American Society for Information Science*, vol. 38, no. 6, pp. 420-442, 1987.
- [595] J. Zhang and R. R. Korfhage, "A distance and angle similarity measure method," *Journal of the American Society for Information Science*, vol. 50, no. 9, pp. 722-778, 1999.
- [596] M. J. McGill, M. Koll and T. Noreault, An Evaluation of Factors Affecting Document Ranking by Information Retrieval Systems, Syracuse, NY: School of Information Systems, Syracuse University, 1979.
- [597] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology*, vol. 26, no. 3, p. 297-302, 1945.
- [598] P. C. Mahalanobis, "On the generalised distance in statistics," *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1, pp. 49-55, 1936.
- [599] D. R. Reising, *Personal Communication*, 2014.
- [600] Y. M. Altman, Accelerating MATLAB Performance: 1001 tips to speed up MATLAB programs, CRC Press, 2014.

Vita

Mr. Trevor J. Bihl graduated from Green Local Schools in Franklin Furnace, Ohio after attending grades 1-12 in the Green Local School system. He graduated with a Bachelor of Science degree in Electrical from Ohio University in Athens, Ohio where he studied various topics, including Electrical Power Systems, Control Theory, Art, Economics, and Sociology. He then began graduate coursework at Ohio University in March of 2005 and ultimately completed a thesis in control theory for a Master's of Science degree in Electrical Engineering in June 2011. Between May 2005 and September 2008 he alternately served as A) an Electronics and Aerospace Engineer at the Air Force Research Laboratory, Space Vehicles Directorate (AFRL/VSSV) at Kirtland AFB, NM for summers 2006 – 2008, B) a graduate research assistant for the General Electric Jet Engine Project at Ohio University for 2005 – 2007, and C) a Stocker Research Fellow at Ohio University in Athens, Ohio for 2007 – 2008. In 2008 he became employed as an Engineer/Scientist by the Schafer Corporation in Dayton, OH and supported spacecraft power system modeling, ballistic missile modeling, and sounding rocket failure analysis. He began supporting the Air Force Institute of Technology (AFIT) Sensor Fusion Lab as a Research Associate in October 2009 and is currently in that position. He holds additional faculty affiliate positions in AFIT's Department of Operational Sciences, Wright State University's (WSU) Department of Biomedical, Industrial, and Human Factors Engineering, and WSU's Department of Pharmacology and Toxicology.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>				
1. REPORT DATE (DD-MM-YYYY) 07-12-2015		2. REPORT TYPE Doctoral Dissertation		3. DATES COVERED (From — To) July 2011 – Dec 2015
4. TITLE AND SUBTITLE Feature Selection and Classifier Development for Radio Frequency Device Identification			5a. CONTRACT NUMBER	
			5b. GRANT NUMBER	
			5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Bihl, Trevor J., Civilian			5d. PROJECT NUMBER JON #15G190F	
			5e. TASK NUMBER	
			5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPAFB OH 45433-7765			8. PERFORMING ORGANIZATION REPORT NUMBER AFIT-ENG-DS-15-D-003	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory, AFMC Attn: AFRL/RVWE (Dr. Vasu Chakravarthy) 2241 Avionics Circle, Bldg 620 WPAFB OH 45433-7734 DSN 798-8269, COMM 937-528-8269 Email: vasu.chakravarthy@us.af.mil			10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RVWE	
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A: APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.				
13. SUPPLEMENTARY NOTES This material has been declared a work of the U.S. Government and is not subject to copyright protection in the United States.				
14. ABSTRACT The proliferation of simple and low-cost devices, such as IEEE 802.15.4 "ZigBee" and Z-Wave, in Critical Infrastructure (CI) increases security concerns. Radio Frequency "Distinct Native Attribute" (RF-DNA) Fingerprinting facilitates biometric-like identification of electronic devices emissions from variances in device hardware. Developing reliable classifier models using RF-DNA fingerprints is thus important for device discrimination to enable reliable Device Classification (a one-to-many looks "most like" assessment) and Device ID Verification (a one-to-one looks "how much like" assessment). AFIT's prior RF-DNA work focused on Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) and Generalized Relevance Learning Vector Quantized Improved (GRLVQI) classifiers. This work 1) introduces a new GRLVQI-Distance (GRLVQI-D) classifier that extends prior GRLVQI work by supporting alternative distance measures, 2) formalizes a framework for selecting competing distance measures for GRLVQI-D, 3) introducing response surface methods for optimizing GRLVQI and GRLVQI-D algorithm settings, 4) develops an MDA-based Loadings Fusion (MLF) Dimensional Reduction Analysis (DRA) method for improved classifier-based feature selection, 5) introduces the F-test as a DRA method for RF-DNA fingerprints, 6) provides a phenomenological understanding of test statistics and p-values, with KS-test and F-test statistic values being superior to p-values for DRA, and 7) introduces quantitative dimensionality assessment methods for DRA subset selection.				
15. SUBJECT TERMS Device Discrimination, Machine Learning, Distinct Native Attribute, Network Security, Pattern Recognition				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT U	18. NUMBER OF PAGES 362
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U		
			19a. NAME OF RESPONSIBLE PERSON Dr. Michael A. Temple	
			19b. TELEPHONE NUMBER (Include Area Code) (937) 255-3636, X4279; michael.temple@afit.edu	